The Dissertation Committee for Abhishek Bhowmick
certifies that this is the approved version of the following dissertation:

Algebraic and analytic techniques in coding theory

Committee:

David Zuckerman, Supervisor

Chandrajit Bajaj

Anna Gal

Shachar Lovett

Eric Price

Algebraic and analytic techniques in coding theory

by

Abhishek Bhowmick, B. Tech.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2015

Dedicated to my parents.

Acknowledgments

A lot of people have made this dissertation possible. The results here have been a consequence of fruitful discussions with various researchers I have interacted with. First, I would like to thank David Zuckerman. I am very fortunate to have him as my supervisor. He had a big impact on me throughout graduate school. Without him, this dissertation would not have been possible. He gave me the confidence to tackle bold problems in my research. During times when I was not making much progress, his encouragement and motivation kept guiding me forward. He has a great approach to research and life in general and I hope to emulate it in future.

Apart from David, there have been many other researchers who have played the role of mentor to me. I was surprised at the amount of time they were willing to put to guide me and I am very lucky to be around such people. They have all given me time to hear out my primitive ideas very patiently and improved my way of thinking. First, I am grateful to Shachar Lovett who has also mentored me during my time in graduate school. I have a learnt a lot from him. His fearless approach to tackling very hard problems is remarkable. It has influenced my style of approaching

research and I am thankful to him for that. I would like to thank Cynthia Dwork for her guidance and advice. I met her in my first year of graduate school when she hosted me at Microsoft Research for the summer. Ever since then, she has been a constant source of motivation and encouragement to me. In my second year, I was fortunate to be hosted by Sanjeev Arora at Princeton University. There, I was introduced to Zeev Dvir. I learnt a lot from Zeev as we worked on some interesting problems together. I am also thankful to Avi Wigderson and Ran Raz for arranging my talk at the Institute for Advanced Study in '12 and in '15.

I am thankful to my hosts at summer internships. I am grateful to Cynthia Dwork, Parikshit Gopalan and Sergey Yekhanin for hosting me in summer '11 and summer '13 at Microsoft Research. I would also like to thank Shachar Lovett for hosting me at UC San Diego in summer '14 and Madhur Tulsiani at TTI-Chicago in summer '15.

Within the department, Anna Gal has been very supportive of my research and I had many fruitful discussions with her during my time in Austin. I also want to thank Chandrajit Bajaj for being like a guide to me. I have learnt a lot from him and thank him for the many hours we spent brainstorming research ideas. I would also like to thank Adam Klivans, Greg Plaxton, Eric Price, Vijaya Ramachandran, Pradeep Ravikumar, Sriram Vishwanath, Felipe Voloch and Brent Waters for their advice and the fruitful discussions I had with them. I am thankful to Lorenzo Alvisi for his encouragement and support and to Lydia Griffith, for her immense help and for making the progress through grad school requirements a breeze.

I thank Madhu Sudan and Ankur Moitra for their valuable suggestions and

encouragement when I visited MIT. I also thank Aditya Bhaskara, Arnab Bhattacharyya, Anindya De, Pooya Hatami, Xin Li, Raghu Meka, Praneeth Netrapalli, Anup Rao, Ankit Rawat, Noga Ron-Zewi, Sushant Sachdeva, Srikanth Srinivasan, Abhradeep Thakurta, Arvindan Vijayraghavan and many others. Also, a very big thanks to the students at the department who made my stay here very enjoyable.

Finally, I would like to thank my parents whose unwavering belief in my abilities kept me going. I am grateful to Komal, Amrita and Ankit for standing by me and their support has been invaluable to me.

Abhishek Bhowmick

December, 2015.

Algebraic and analytic techniques in coding theory

Abhishek Bhowmick, Ph.D.
The University of Texas at Austin, 2015

SUPERVISOR: David Zuckerman

Error correcting codes are designed to tackle the problem of reliable transmission of data through noisy channels. A major challenge in coding theory is to efficiently recover the original message even when many symbols of the received data have been corrupted. This is called the *unique decoding problem* of error correcting codes. More precisely, if the user wants to send K bits, the code *stretches* K bits to N bits to tolerate errors in the N bits. Then the goal is to recover the original K bits of the message.

Often, the receiver requires only a certain part of the message. In such cases, analysing the entire received data (word) becomes prohibitive. The challenge is to design a *local decoder* which queries only few locations of the received word and outputs the part of the message required. This is known as *local decoding* of an error correcting code.

The unique decoding problem faces a certain combinatorial barrier. That is, there is a limit to the number of errors it can tolerate in order to uniquely identify the correct message. This is called the unique decoding radius. A major open problem

is to understand what happens if one allows for errors beyond this threshold. The goal is to design an algorithm that can recover the right message, or possibly a list of messages (preferably a small number). This is referred to as *list decoding* of an error correcting code.

At the core of many such codes lies polynomials. Polynomials play a fundamental role in computer science with important applications in algorithm design, complexity theory, pseudo-randomness and machine learning.

In this dissertation, we improve our understanding of well known classes of codes and discover various properties of polynomials. As an additional consequence, we obtain results in a suite of problems in effective algebraic geometry, including Hilbert's nullstellensatz, ideal membership problem and counting rational points in a variety.

Table of Contents

Chapter 1

Introduction

1.1 Error correcting codes

Error correcting codes are designed to tackle the problem of robust transmission of data through noisy channels. The theory of coding theory dates back to the seminal work of Shannon [150] and Hamming [92]. The principle involves adding redundancy in a systematic way to the message before transmitting it with the hope that even after some bits get corrupted, the decoder can decode back the original message. There are certain basic notions that we now introduce. We fix our alphabet to be $\{0,1\}$ for now.

- Encoder $E : \{0,1\}^K \to \{0,1\}^N$; the mapping that takes the K input bits and *stretches* it to N bits by adding redundancy.
- Decoder $D : \{0,1\}^N \to \{0,1\}^K$; the mapping that takes the N received bits and decodes back the original message.
- Rate $R = K/N$; A measure of the rate of information sent per unit bit transmitted.

- Hamming distance $\Delta(x, y)$; the number of coordinates i where x_i and y_i differ.
- Minimum distance δ; The minimum pairwise distance (normalized by dividing by n) between any two distinct codewords.

The first goal for a code designer is to ensure that the code has high rate R and high minimum distance δ. The two quantities cannot be increased simultaneously for obvious reasons. There are various bounds which control the behavior of R and δ. The other important goals are the algorithmic constructions of the encoder and the decoder. The encoder E is usually specified in the description of the code. The harder question is therefore the algorithmic decoder D. The goal is on sending m, if y is the noise added to the transmission, then the decoder on seeing the noisy message should still recover m. More precisely, we want $D(E(m) + y) = m$. The level of noise is typically measured by the Hamming weight wt, that is, the number of non zero elements in y^N. If $\text{wt}(y) \leq \lfloor (d-1)/2 \rfloor$, then there is only one codeword $E(m)$ that is $\text{wt}(y)$ far from the received word. However, this does ensure an efficient algorithm to find it. This is the case of unique decoding.

Often, we are interested in only recovering a part of the message, say the first bit. In such cases, looking at all N bits might be too time consuming. Therefore, we require an algorithm that should look at only a few coordinates (say 3) and still recover the first bit of the message with high probability. This leads to the concept of local decodable codes. We will discuss this more in a later subsection.

We will also be interested in the case when $\text{wt}(y) > \lfloor (d-1)/2 \rfloor$. In this case, we try to bound the list of codewords that are candidate encodings of the

right message. Note that we can no longer guarantee a unique codeword at such a large distance. This is called list decoding. The goal here is again twofold. First, one needs to prove a combinatorial bound on the size of the list of codewords that are within a certain distance of the received word. Second, there should be an algorithmic way to output the list of the above codewords. We shall revisit the problem of list decoding shortly.

1.2 Polynomial Codes

An important and well studied family of codes is the family of polynomial based codes. Fix a finite field $\mathbb{F} = \mathbb{F}_q$. Let $d \in \mathbb{N}$. The univariate polynomial based code, commonly called the Reed Solomon code, is defined as follows. The message space consists of degree $\leq d$ polynomials over $\mathbb{F}$ and the codewords are evaluation of these polynomials on $\mathbb{F}$. In more detail, every message is a univariate polynomial $f = \sum_{i=0}^{d} a_i x^i$. The encoder E maps this to $(f(x) : x \in \mathbb{F}) \in \mathbb{F}^q$. Similarly, in the multivariate polynomial based code, commonly called the Reed Muller code, the message space consists of n variate degree $\leq d$ polynomials over $\mathbb{F}$ and the codewords are evaluation of these polynomials on $\mathbb{F}^n$. We call this $\mathrm{RM}_{\mathbb{F}}(n, d)$. The metric in the space of all n-variate functions is defined as follows: For $f, g : \mathbb{F}^n \to \mathbb{F}$, $\delta(f, g) = \mathbf{Pr}[f(x) \neq g(x)]$. Let $\delta_{\mathbb{F}}(d)$ denote the minimum distance of $\mathrm{RM}_{\mathbb{F}}(n, d)$. Let $d = a(q-1) + b$ where $0 \leq b < q - 1$. We have by the Schwartz-Zippel lemma,

$$\delta_{\mathbb{F}}(d) = \frac{1}{q^a}\left(1 - \frac{b}{q}\right).$$

Since polynomial based codes involve fundamental objects, i.e., low degree polynomials, they have found tremendous application in various areas of computer

science. In this dissertation, among other things, we shall answer some open questions related to multivariate polynomials with applications to Reed Muller codes.

1.3 Locally Decodable Codes

A Locally Decodable Code (LDC) is a special kind of code that allows the receiver to decode a *single* symbol of the message by querying a small number of positions in a corrupted encoding. Since the early 1990's, LDC's have found exciting applications in various areas ranging from data transmission to complexity theory to cryptography and privacy. We refer the reader to [166, 180] for more background.

A central research question, which is far from being solved, is understanding the best possible 'stretch' of an LDC with a constant number of queries. That is, how large N has to be as a function of K. There has been a significant amount of work done; see ([105, 70, 112, 174, 175, 16, 178, 59]). The best family of LDCs in the constant query regime are Matching Vector (MV) codes, which have much shorter codeword length than the classical constructions. These were discovered by Yekhanin [178] and Efremenko [59] and analyzed in more detail by Dvir et al ([56]).

Lower Bounds on MV Codes. Our main contribution in this area is a lower bound on the *s*tretch of MV codes [32]. The first result proves a quadratic lower bound ($N = \Omega(K^2)$) which resolves a conjecture raised in [56].

Theorem 1.3.1 (Informal)**.** *Consider an infinite family of $O(1)$-query Matching Vector code $C_n : \mathbb{F}^K \to \mathbb{F}^N$ for $n \in \mathbb{N}$, where $K(n)$ and $N(n)$ are growing functions of n. Then*

$$N \geq k^{2-o(1)}$$

The result holds even when the number of queries is not constant. The second result states that under a well known conjecture from additive combinatorics, one needs a super polynomial stretch ($N = K^{\log\log K}$), thus ruling out efficient MV codes for constant number of queries.

Theorem 1.3.2 (Informal)**.** *Assume the PFR conjecture (Conjecture 1) holds. Consider an infinite family of $O(1)$-query Matching Vector code $C_n : \mathbb{F}_q^K \to \mathbb{F}_q^N$ for $n \in \mathbb{N}$, where $K(n)$ and $N(n)$ are growing functions of n. For large enough n, Then*

$$N = \exp\left(\Omega_m\left(\log k \ \ \log\log k\right)\right)$$

See Chapter 2 for details. This is joint work with Zeev Dvir and Shachar Lovett.

Upper Bounds on MV Codes. The best constructions of MV codes come from constructions of a particular type of polynomial called an OR polynomial. Smaller degree OR polynomials lead to better MV codes. We show a barrier why existing techniques have failed to improve the degree bounds of OR polynomials for more than a decade. In fact, the barrier we discover applies to many other fundamental problems in complexity theory [34]. See Chapter 3 for details. This is joint work with Shachar Lovett.

1.4 List Decodable Codes

The concept of *list decoding* was introduced by Elias [60] and Wozencraft [176] to decode *error correcting codes* beyond half the minimum distance. The objective of list decoding is to output all the codewords within a specified radius around the

received word. After the seminal results of Goldreich and Levin [67] and Sudan [153] giving list decoding algorithms for the Hadamard code and the Reed-Solomon code respectively, there has been tremendous progress in designing list decodable codes. See the excellent surveys of Guruswami [89, 88] and Sudan [154].

List decoding has applications in many areas of computer science including hardness amplification in complexity theory [155, 165], construction of hard core predicates from one way functions [67, 2], construction of extractors and pseudorandom generators [168, 157, 149, 91] and computational learning [116, 101].

List Decoding Radius of Reed Muller Codes. Despite so much progress, the largest radius up to which list decoding is tractable (the *list decoding radius*) is still a fundamental open problem even for well studied codes like Reed-Solomon (univariate polynomials) and Reed-Muller codes (multivariate polynomials). The goal of this work is to analyze Reed-Muller codes over small fields $\mathbb{F}$ and small degree d. The list decoding radius was conjectured to *approach* the minimum distance of the code. See Chapter 4 for a precise definition. This was proved for the $d = 1$ case [67, 68], the $\mathbb{F} = \mathbb{F}_2$ case [74] and the $d = 2$ case [73]. It was conjectured [74] to be true for all fixed fields and fixed degree. Our main contribution is a positive resolution of the above conjecture for fixed prime fields [35]. That is, we prove that the list decoding radius of Reed Muller codes equals the minimum distance of the code for fixed prime fields and fixed degree. Moreover, we give a tight bound on the weight distribution of generalized Reed Muller codes over prime fields. This is a fundamental problem in coding theory; see Research Problem (15.1) in [124]. We skip the weight distribution statement below. For details, see Chapter 4. More precisely, we have the following.

Let $\mathcal{P}_d(\mathbb{F}^n)$ denote the class of degree $\leq d$ polynomials $f : \mathbb{F}^n \to \mathbb{F}$. Let $\delta_{\mathbb{F}}(d)$ denote its minimum distance. For $\mathrm{RM}_{\mathbb{F}}(n, d)$, $\eta > 0$, let

$$\ell_{\mathbb{F}}(n, d, \eta) := \max_{g:\mathbb{F}^n \to \mathbb{F}} |\{f \in \mathcal{P}_d(\mathbb{F}^n) : \delta(f, g) \leq \eta\}| .$$

Theorem 1.4.1 (Informal)**.** *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta_{\mathbb{F}}(d) - \varepsilon) \leq c_{p,d,\varepsilon}.$$

In follow up work, we extend this to large fields and prove that the list decoding radius equals the minimum distance for fixed degree and all prime fields [36]. This is a consequence of a theorem about pseudorandomness of polynomials which is mentioned in more detail in the next section.

Applications of List Decoding to Randomness Extraction. High-quality randomness is needed for a variety of applications. However, most physical sources are only weakly random. Moreover, such weak sources arise in cryptography when an adversary learns information about a uniformly random string. It is therefore natural and important to try to extract the usable randomness from a weak source. It is impossible to extract even one bit of randomness from a natural yet large enough class of sources using a single function [143]. There are two ways to counter this. One is to extract with the help of a small amount of randomness; this is called a seeded extractor [131]. Our focus is on the second way: to extract only from more structured sources (and not allow any auxiliary randomness). Such a function is called a *deterministic* (or seedless) extractor. Our main contribution is

to extract randomness from a very general class of *additive* sources which satisfy a certain list decodability property [33]. Our work generalizes many existing results [104, 65, 42, 177, 117, 63, 55]. For details, see Chapter 7. This is joint work with Ariel Gabizon, Le Thai Hoang and David Zuckerman.

1.5 Polynomials and computation

Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $f : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree d. We say f is *equidistributed* if f takes every value in $\mathbb{F}$ with the same frequency, else it is *biased*. We say that f is *low rank* if it can be expressed as a composition of bounded number of lower degree polynomials and high rank otherwise.

Definition 1.5.1 (bias)**.** *The* bias *of a function $f : \mathbb{F}^n \to \mathbb{F}$ is defined to be*

$$\mathrm{bias}(f) = \mathbb{E}_{x \in \mathbb{F}^n}[e(f(X))],$$

where $e(x) := e^{2\pi i x}$.

Definition 1.5.2 (rank)**.** *Let $d \in \mathbb{N}$ and $f : \mathbb{F}^n \to \mathbb{F}$. Then $\mathrm{rank}_d(f)$ is defined as the smallest integer r such that there exist polynomials $h_1, \ldots, h_r : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d-1$ and a function $\Gamma : \mathbb{F}^r \to \mathbb{F}$ such that $f(x) = \Gamma(h_1(x), \ldots, h_r(x))$. If $d = 1$, then the rank is 0 if f is a constant function and is ∞ otherwise. If f is a polynomial, then $\mathrm{rank}(f) = \mathrm{rank}_d(f)$ where $d = \deg(f)$.*

Green and Tao [80] and Kaufman and Lovett [108] proved that if a low degree polynomial has high bias, then it has low rank. However, a crucial restriction they have is that the underlying field is a fixed prime.

We generalize this result to all prime fields (possibly growing with n) in Chapter 5 [36] in joint work with Shachar Lovett. In some cases, we extend this to nonprime fields in Chapter 6 [25] in joint work with Arnab Bhattacharyya. We further give some consequences of this theorem in list decoding of Reed Muller codes and effective algebraic geometry.

1.5.1 Application to list decoding RM codes

First, we prove that the list decoding radius equals the minimum distance for fixed degree and all prime fields [36] and in some cases, also for nonprime fields [25]. For ease of exposition, assume $d < p$, that is, the degree is smaller than the field size. It is well known from the seminal results of Arora and Sudan [5] and Sudan, Trevisan and Vadhan [155] that Reed Muller codes are list decodable from an error rate of $1 - \sqrt{d/p}$ using ideas of Guruswami and Sudan [90]. This is commonly known as the Johnson radius. There are very few classes of codes that are known to be list decodable beyond the Johnson radius.

We prove that Reed Muller codes can be list decodable from error rate approaching $1 - d/p$, which is the minimum distance and is higher than the Johnson radius. Also, it is well known that this is tight.

Next, we highlight some consequences in effective algebraic geometry.

1.5.2 Application to problems in effective algebraic geometry

1. A finite field Hilbert nullstellensatz.

Hilbert's strong nullstellensatz establishes a relationship between algebra and geometry and is a fundamental theorem in algebraic geometry. It states the following: given a collection of polynomials, if f vanishes on the set of common zeroes of the polynomials, then some power of f lies in the ideal generated by the collection of polynomials. The area of effective nullstellensatz tries to bound two quantities. First, it bounds what power f should be raised to for the theorem to hold. And second, it bounds the degrees of the coefficient polynomials when representing a power of f as a member of the ideal. We prove effective versions of these bounds when all polynomials are of fixed degree. Under the regime of fixed degree, we are able to achieve something much stronger, which we highlight shortly. We call this the finite field analogue of the Hilbert nullstellensatz.

Theorem 1.5.3 (Finite field Hilbert Nullstellensatz)**.** *Let $c, d \in \mathbb{N}$. Let $P_1, \ldots P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Assume that $Q(x) = 0$ whenever $P_1(x) = \ldots = P_c(x) = 0$. Then there exist $R_1, \ldots R_c \in \mathcal{P}_D(\mathbb{F}^n)$, $D^{(6.1.3)} = p.O_{d,c}(1)$, such that*

$$Q(x) \equiv \sum_{i=1}^{c} R_i(x) P_i(x).$$

The improvement from the usual nullstellensatz for closed fields here is two-fold:

- There is no exponent in $Q(x)$ which makes this a stronger conclusion.
- $Q(x)$ has to vanish only on the common zeros of the P_i's in the finite field and not its closure which makes this a weaker requirement.

Stated as a Hilbert nullstellensatz result, we have the following.

Theorem 1.5.4 (Effective Hilbert Nullstellensatz)**.** *Let $c, d \in \mathbb{N}$. Let $P_1, \ldots P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Assume that $Q(x) = 0$ whenever $P_1(x) = \ldots = P_c(x) = 0$ for x* ***in the algebraic closure of*** *$\mathbb{F}$. Then there exist $R_1, \ldots R_c \in \mathcal{P}_D(\mathbb{F}^n)$, $D^{(6.1.3)} = p.O_{d,c}(1)$ and $\mathbf{r} = \mathbf{1}$, such that*

$$Q(x)^r \equiv \sum_{i=1}^{c} R_i(x) P_i(x).$$

In this work, we focus on the setting of constant d and c. The first effective result in this direction was due to Hermann [95] in 1926 who proved $D = d^{O(2^n)}$ and $r = O_{d,c}(1)$. In 1987, Brownawell [45] and later Kollar [113] in 1988 in breakthrough results, proved a singly exponential bound in n. In fact, for constant c, then achieve $D = O(d^c)$ and $r = O_{d,c}(1)$. Surprisingly, Green and Tao [80] obtained $r = 1$ but $D = O_{d,c,p}(1)$. Note that they have an Ackermann dependence on the field size. However, $r = 1$ immediately leads to a fast ideal membership algorithm. We obtain $D = p.O_{d,c}(1)$ and $r = 1$. Note that we made the dependence on p linear from Ackermann.

2. Ideal membership.

A related problem is that of ideal membership. Here the problem is given a collection of polynomials, and a polynomial f, find if f belongs to the ideal generated by the above collection. The challenge is to do this using an efficient algorithm. We prove the following algorithmic result.

Theorem 1.5.5 (Algorithmic Ideal Membership)**.** *Let $c, d \in \mathbb{N}$. Let $P_1, \ldots, P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Let $I = \langle P_1, \ldots, P_c \rangle$. Then we have an algorithm that performs $n^{p.O_{d,c}(1)}$*

field operations and decides if $Q \in I$.

The ideal membership problem is known to be EXPSPACE-hard over the rationals. Over finite fields, the fastest general algorithm was by Buchberger [46] where we formally constructed Gröbner bases in a series of works, with running time d^{2^n}. Green and Tao [80] gave an algorithm that was polynomial in n with everything else fixed, that is a running time $n^{O_{d,c,p}(1)}$. The dependence on p was Ackermann however. We provide an algorithm with running time $n^{p.O_{d,c}(1)}$. Again, note that the exponent of n was improved from Ackermann to linear in p.

3. Counting points in rational varieties.

Finally, we come to the problem of counting the number of rational points in a variety. The exact problem of detection of rational points is NP hard; see for example [1, 98, 115, 97, 66, 75].

Given polynomials $f_1, \ldots f_c : \mathbb{F}^n \to \mathbb{F}$, let $V(f_1, \ldots, f_c) \subseteq \mathbb{F}^n$ denote the set of common zeroes of f_i's.

Lemma 1.5.6 (Rational points in varieties)**.** *Let $c, d, t, u \in \mathbb{N}$. Let $P_1, \ldots, P_c \in \mathcal{P}_d(\mathbb{F}^n)$. There is a randomized algorithm that performs $O_{d,c,t,u}(n^d) + |\mathbb{F}|^{O_{d,c,t}(1)}$ field operations and performs the following with probability $1 - \frac{1}{|\mathbb{F}|^t}$:*

1. *Decide if $V_p(P_1, \ldots, P_c)$ is empty.*

2. *Output an integer N such that $N = (1 \pm |\mathbb{F}|^{-u})|V_p(P_1, \ldots, P_c)|$.*

4. Holes in the number of rational points.

The proof of above lemma also implies that the number of common zeroes does not span all possible values. They only lie in the following union of intervals

$$\bigcup_{i=1}^{|\mathbb{F}|^{c'}} \left[i \cdot |\mathbb{F}|^{n-c'} \left(1 - |\mathbb{F}|^{-u}\right), i \cdot |\mathbb{F}|^{n-c'} \left(1 + \mathbb{F}|^{-u}\right) \right].$$

As a special case, we have a strengthening of the Chevellay-Warning theorem in the setting of fixed c, d. The Chevellay-Warning theorem states that if a collection $P_1, \ldots, P_c \in \mathcal{P}_d(\mathbb{F}^n)$ with $dc < n$) has one common solution, it has at least $|\mathbb{F}|$ many solutions.

Corollary 1.5.7. *Let $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ be polynomials of degree d. Then, if the collection has at least one solution, then it has at least $|\mathbb{F}|^{n-O_{c,d}(1)}$ many solutions.*

Such a result for constant size prime fields was proved by Lovett in [121]. It is noteworthy to compare the above bound with the Az-Katz theorem [6, 106], which says that the number of solutions is at least $|\mathbb{F}|^{n/d-c}$. More formally,

Theorem 1.5.8 (Ax-Katz theorem)**.** *Let $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ be polynomials of degree d. Then, if the collection has at least one solution, then it has at least $|\mathbb{F}|^{n/d-c}$ solutions.*

The rest of the document is organized as follows: Chapter 2 contains results on lower bounds of MV codes. Chapter 3 contains results on the barrier in complexity theory problems. Chapter 4 shows results on the list decoding radius of RM codes for fixed primes, Chapter 5 for fixed nonprimes and Chapter 6 for large prime

fields and applications in effective algebraic geometry. Chapter 7 contains results on the application of list decoding to randomness extraction from structured sources.

Chapter 2

Lower Bounds on MV codes

2.1 Introduction

Recall that locally decodable codes (LDCs) allow the receiver to decode a *single* symbol of the message by querying a small number of positions in a corrupted encoding. More formally, an (q, δ, ϵ)-LDC encodes K-symbol messages x to N-symbol codewords $C(x)$, such that for every $i \in [K]$, the symbol x_i can be recovered with probability $1 - \epsilon$, by a randomized decoding procedure that makes only q queries, even if the codeword $C(x)$ is corrupted in up to δN locations. Since the early 90's, LDC's have found exciting applications in various areas ranging from data transmission to complexity theory to cryptography/privacy. We refer the reader to [166, 180] for more background.

A central research question, which is far from being solved, has to do with understanding the best possible 'stretch' of an LDC with a constant number of queries. That is, how large N has to be as a function of K for constant q and with constant δ, ε (these two last parameters are not our focus here and we will generally assume them to be small fixed constants). For $q = 1, 2$ this question is completely

answered. There are no LDC's for $r = 1$ [105] and the best LDC's with $q = 2$ have exponential encoding length [70, 112]. For $q > 2$ there are huge gaps in our understanding. Katz and Trevisan were the first to study this problem [105] and, today, the best general lower bounds on N are slightly super-linear bounds of the form $\tilde{\Omega}\left(K^{1+1/(\lceil r/2\rceil-1)}\right)$ [174]. Notice that, when the number of queries is 3 or 4, these bounds are quadratic (see also [112, 175] for the $q = 3, 4$ case). The upper bounds were, until recently, those coming from polynomial codes and were of the order of $N \leq \exp\left(K^{\frac{1}{q-1}}\right)$. Improved upper bounds, breaking this barrier slightly, were given in [16].

This state of affairs changed dramatically when, in a breakthrough paper, Yekhanin [178] developed a new approach for constructing LDCs, called Matching Vector codes (MV codes), that have much shorter codeword length than polynomial codes. Efremenko [59] was the first to show that this approach could yield codes with subexponential encoding length (Yekhanin's paper showed this under a number theoretic assumption). More refinements and improvements to this new framework were obtained [134, 111, 100, 125, 56, 17] to give LDC's with q queries and with encoding length that grows, when q is a constant, roughly like

$$N \sim \exp\exp\left((\log K)^{O(1/\log q)}(\log\log K)^{1-1/\log q}\right).$$

While significantly smaller than the length of polynomial codes, the codeword length of these new codes is still super polynomial in K. The most general setting of parameters was addressed in [56] where the authors had given a black box construction of q query MV codes using $q-$*restricted* MV families in $\mathbb{Z}_m^n$. At the

heart of an MV code is a Matching Vector family (MV family). Using the standard definition of MV families, this implied m query MV codes using MV families in Z_m^n. In this basic, yet general reduction, it was shown that upper bounds on MV families would lead to lower bounds on the encoding length of MV codes. With this motivation in mind, the authors in [56] made a conjecture on the upper bound on the size of MV families which would lead to lower bounds on the encoding length of MV codes under the basic framework. We note that Yekhanin in [178] used restricted MV families in $\mathbb{Z}_p^n$ where p is a very large Mersenne prime and used a specialized technique to reduce the number of queries from p to 3. Another instance of reduction in the number of queries from what the standard construction gives, was given by Efremenko [59] where he again used restricted MV families. A certain gadget was discovered using computer search whereby the author worked in $\mathbb{Z}_{511}$ but got down the number of queries to 3 from the basic bound of 511.

A Matching Vector Family (MV Family) is a combinatorial object that arises in several contexts including Ramsey graphs, weak representation of OR polynomials and recently in constant query locally decodable codes (LDCs). It is defined by two ordered lists $U = (u_1, \cdots u_t)$ and $V = (v_1, \cdots v_t)$ where $u_i, v_j \in \mathbb{Z}_m^n$ and m and n are integers greater than 1. The property that the two lists have to satisfy is the following: for all $i \in [t]$, $\langle u_i, v_i \rangle = 0 \ (mod\ m)$ whereas for all $i \neq j \in [t]$, $\langle u_i, v_j \rangle \neq 0 \ (mod\ m)$. By $\langle \cdot, \cdot \rangle$ we denote the standard inner product. Let us call this the standard definition of a MV family. If in addition, all the inner products $\langle u_i, v_j \rangle \ (mod\ m)$ lie in a set of size q, then it is called a $q - restricted$ MV family. Note that $q = m$ corresponds to the standard MV family. The size of the MV family

is t, the number of vectors in the list. In this paper, we shall prove upper bounds on $q-restricted$ MV families in the first part and on standard MV families in the later part.

Before we state the bounds for MV families, we give a quick description of how to construct an MV code from an MV family. Let $U = \{u_1, \ldots, u_K\}$ and $V = \{v_1, \ldots, v_K\}$ form an MV family in $\mathbb{Z}_m^n$. Let $m|q-1$. Let ω be a primitive m-th root of unity in $\mathbb{F}_q$. The encoder $E : \mathbb{F}_q^K \to \mathbb{F}_q^N$, where $N = m^n$ is defined as follows. Given $a = a_1, \ldots, a_K$, $E(a) = \left(\sum_{i=1}^K a_i \omega^{\langle u_i, x\rangle} : x \in \mathbb{Z}_m^n\right)$.

Now, to decode say a_i, assume the received word is error free as by a union bound over the number of queries we can assume this is indeed the case with high probability. Now, query the received word at coordinates $S = \{x + \ell v_i : 0 \leq j \leq m-1\}$ and add them. Indeed, it can be shown that the sum of those coordinates exactly equals a_i.

Let $\mathbf{MV}(m,n)$ denote the largest t such that there exists a MV family of size t in $\mathbb{Z}_m^n$. Analogously, let $\mathbf{MV}(m,n,q)$ denote the largest t such that there exists a $q-restricted$ MV family of size t in $\mathbb{Z}_m^n$. The question of bounding $\mathbf{MV}(m,n)$ (or $\mathbf{MV}(m,n,q)$) is closely related to the well-known combinatorial problem of set systems with restricted modular intersections [9, 147, 86, 87] (in this setting the vectors u_i, v_i are required to have entries that are either 0 or 1). The systematic study of this more general problem, in the context of MV codes, was initiated in [56]. The setting of prime m is well understood. For large prime $m = p$, it is known that $\mathbf{MV}(p,n) = O\left(p^{n/2}\right)$ [56]. Infact, this is almost tight. When m is a small prime, again we have a tight upper bound of $O\left(n^{p-1}\right)$ [9]. Surprisingly, the

setting of small composite m leads to very useful constructions of Ramsey graphs and constant query LDCs. This is due to a construction of MV family over Z_6 by Grolmusz [86] of superpolynomial size in contrast to a polynomial upper bound when m is a small prime. Thus, it is interesting to study the behavior of MV families for small composite m, and more generally arbitrary general composites. We will see later the connection between upper bounds on $\mathbf{MV}(m, n, q)$ and lower bounds on the encoding lengths of MV Codes (a family of LDCs). For general m, the best upper bound known [56] is $\mathbf{MV}(m, n) \leq m^{n-1+o_m(1)}$, with $o_m(1)$ denoting a function that goes to zero when m grows. It was conjectured in [56] that an upper bound of $\sim m^{n/2}$ should hold for any m (not just prime). This would be tight for large m as there are constructions of MV families almost meeting this bound [182]. However, the proof method used in [56] to prove the $O\left(p^{n/2}\right)$ bound does not extend to non primes. In this work, we prove the conjecture for $q-restricted$ MV families in $\mathbb{Z}_m^n$, for any m as long as $q < \max\{\frac{o(n)\log m}{\log(o(n)\log m)}, 2^{o(n)}\}$ (See Theorem 1). When $m = p$ is a fixed small prime, it follows from [9] that $\mathbf{MV}(p, n) = O\left(n^{p-1}\right)$. On the other hand, when m is a fixed composite, say $m = 6$, there exists a MV family of superpolynomial size $\Omega\left(\exp\left(\log^2 n/\log\log n\right)\right)$ [86]. We prove a stronger upper bound on $\mathbf{MV}(m, n)$, compared to Theorem 1 in such a case assuming a well known conjecture in additive combinatorics (see Theorem 2). Table 2.1 lists the known and new upper bounds on MV families.

Theorem 1. *For all $m \geq 2, n \geq 1$ we have*

$$\mathbf{MV}(m, n, q) \leq \min\{q^{O(q\log q)}m^{n/2}, q^{O(\log m)}m^{n/2}\}$$

m	$\mathbf{MV}(m,n)$ or $\mathbf{MV}(m,n,q)$
general prime	$\mathbf{MV}(m,n) \leq O\left(m^{n/2}\right)$ [56]
general composite	$\mathbf{MV}(m,n,q) \leq \min\{q^{O(q\log q)}m^{n/2}, q^{O(\log m)}m^{n/2}\}$ (Theorem 1)
small, fixed prime	$\mathbf{MV}(m,n) \leq O\left(n^{m-1}\right)$ [9]
small, fixed composite	$\mathbf{MV}(m,n) \leq 2^{O_m(n/\log n)}$ (Theorem 2 under Conjecture 1)

Table 2.1: List of upper bounds on $\mathbf{MV}(m,n)$, $\mathbf{MV}(m,n,q)$

Hence, Theorem 1 resolves the conjecture of [56] for any m, such that $q < \max\{\frac{o(n)\log m}{\log(o(n)\log m)}, 2^{o(n)}\}$. In fact, if q is a constant, the first bound is of the form $\mathbf{MV}(m,n,q) \leq cm^{n/2}$ for some constant c. When $m >> n$, our bound is quite close to the best known construction of MV families which gives $\mathbf{MV}(m,n) \geq \left(\frac{m+1}{n-2}\right)^{n/2-1}$ [182].

Our second result assumes the polynomial Freiman-Ruzsa conjecture (PFR) conjecture (discussed below) and gives a stronger upper bound on the size of MV families when m is a constant and n grows.

Before we state the conjecture, we need to define what a difference set is. For an abelian group G let $A \subseteq G$. Then the difference set

$$A - A = \{a_1 - a_2 : a_1, a_2 \in A\}$$

The PFR conjecture has been stated in its complete generality in [140]. We rely on a special case of it stated below.

Conjecture 1 (PFR Conjecture in $\mathbb{Z}_m^n$)**.** *Suppose $A \subseteq \mathbb{Z}_m^n$ and $|A - A| \leq \lambda \cdot |A|$. Then one can find a subgroup H of size at most $|A|$ such that A can be covered by $\lambda' = \lambda^{c_m}$ many translates of H, where c_m depends only on m.*

We note that the PFR conjecture has already found several applications in computer science. Ben-Sasson and Zewi [21] used it to construct two-source extractors from affine extractors; and Ben-Sasson, Lovett and Zewi [20] used it to bound the deterministic communication complexity of functions whose corresponding matrix has low rank. Our work provides another application for the PFR and demonstrates its wide-reaching applicability. We further note that a quasi-polynomial version of the PFR conjecture was recently proved by Sanders [142] (see also the exposition in [122]). Unfortunately, all the applications discussed above require the truly polynomial version of the conjecture, and so cannot apply to Sanders' result.

We now state the second theorem.

Theorem 2. *Assuming the PFR conjecture over $\mathbb{Z}_m^n$ (Conjecture 1) we have*

$$\mathbf{MV}(m,n) \leq \exp\left(c(m)\frac{n}{\log n}\right),$$

with $c(m)$ an explicit function of m.

From a technical point of view, one of the ingredients in this work builds on the recent work of Ben-Sasson, Lovett and Zewi [20] who used the PFR conjecture to show that matrices over $\mathbb{Z}_2$ with large bias (say, with many more ones than zeros) and small rank must contain a large monochromatic sub-matrix. An important ingredient in our proof is a generalization of their results from $\mathbb{Z}_2$ to $\mathbb{Z}_m$ for all m, not necessarily prime. We note however that this is just one ingredient in our overall proof.

2.1.1 Lower Bounds on LDCs

The following is a corollary of Theorem 1.

Corollary 3. *For an arbitrary positive integer m, consider an infinite family of q-query Matching Vector code $C_n : \mathbb{F}^k \to \mathbb{F}^N$ for $n \in \mathbb{N}$, where $k(n)$ and $N(n)$ are growing functions of n, constructed using the black box reduction from a q-restricted Matching Vector Family in $\mathbb{Z}_m^n$ ([56]). For large enough n, if $q < \max\{\frac{o(n)\log m}{\log(o(n)\log m)}, 2^{o(n)}\}$, then*

$$N \geq k^{2-o(1)}$$

Specifically, if $q = O(1)$, then $N = \Omega\left(k^2\right)$.

Next we have the following corollary from Theorem 2.

Corollary 4. *For some arbitrary positive integer m, assume the PFR conjecture over $\mathbb{Z}_m^n$ (Conjecture 1). Consider an infinite family of m-query Matching Vector code $C_n : \mathbb{F}_q^k \to \mathbb{F}_q^N$ for $n \in \mathbb{N}$, where $k(n)$ and $N(n)$ are growing functions of n, constructed using the black box reduction from a standard Matching Vector Family in $\mathbb{Z}_m^n$ ([56]). For large enough n, if $m = O(1)$, then*

$$N = \exp\left(\Omega_m\left(\log k \ \ \log\log k\right)\right)$$

Thus Corollary 4 states that, assuming Conjecture 1, MV codes with constant number of queries must have super polynomial encoding length in the basic framework. Note that we get the same bound in Efremenko's framework for 3 queries. This is because the form of the superpolynomial bound is assuming a constant m

and applying our bound to Efremenko's work again leads to a superpolynomial bound as $m = 511$ in his setting (another constant). (He uses $\mathbb{Z}_{511}$ to construct the MV family and further reduces the number of queries to 3.) This essentially means that in order to construct polynomial length codes, one needs to construct MV families in $\mathbb{Z}_m^n$ for non-constant m and use some specialized gadget to reduce the number of queries. One way is to ensure it is a $q - restricted$ (constant q) MV family. This automatically ensures q query decoding. However, the quadratic lower bound continues to hold even in this scenario for constant q. To beat the quadratic lower bound for constant query MV codes, one needs to construct $q - restricted$ MV families for growing m and q large enough to break the upper bound and then develop some special gadget to get the number of queries down further from q to some constant.

2.1.2 Proof Overview

The proof of Theorem 1 relies on intuitions coming from the theory of two-source extractors [48], which are functions of two variables $F(X, Y)$ such that the output of F is distributed in a close-to-uniform fashion whenever the two inputs are drawn, independently, from two distributions of sufficiently high entropy. Since our proof does not use two-source extractors explicitly we do not define them formally and just use them to explain the high level idea behind the proof. It is a well known fact [48] that the inner product function $F(X, Y) = \langle X \rangle Y$, say over $\mathbb{Z}_2^n \times \mathbb{Z}_2^n$ is a good two source extractor when the two inputs X and Y are both drawn uniformly from sets $S_X, S_Y \subseteq \mathbb{Z}_2^n$ of size larger $2^{n/2}$. This immediately suggests a connection to MV families, since, if we take $S_X = U$ and $S_Y = V$ for a MV family U, V in $\mathbb{Z}_2^n$,

we would get a completely non-uniform output (it will be zero with exponentially small probability). This means that the size of U, V is bounded from above by approximately $2^{n/2}$.

If we try to use a similar argument over $\mathbb{Z}_m$ we run into trouble since the inner product function modulo m is *not* a good two source extractors for sources of size $m^{n/2}$. Take, for example, $S_X = S_Y = \{0, 2, 4\}^n \subseteq \mathbb{Z}_6^n$ and observe that $\langle X \rangle Y$ is always divisible by 2 and so is far from being uniformly distributed over $\mathbb{Z}_6$. It is, however, possible to show that this example is, in some sense, the only example and that, in general, we can always find a certain number of elements of either S_X or S_Y that 'agree' modulo some factor of m. This observation suggests proving Theorem 1 by induction on the number of factors of m, which is the way we proceed.

The proof of Theorem 2 uses a slightly different view of MV families as matrices with certain zero/non-zero pattern and small rank. Specifically, for a MV family U, V of size t in $\mathbb{Z}_m^n$ consider the $t \times t$ matrix P whose (i, j)'th entry is $\langle u_i \rangle v_j \mod m$. The definition of a MV family implies that P has zeros on the diagonal and non-zeros everywhere else. If m was a prime, we could think of $\mathbb{Z}_m$ as a field $\mathbb{F}$ and say that, since P is the inner product matrix of vectors of length n over a field, it must have rank at most n. Conversely, every $t \times t$ matrix over a field $\mathbb{F}$ with these properties (zero on the diagonal and non-zero off the diagonal) and with rank n gives a MV family of size t in $\mathbb{F}^n$.

Assume for the purpose of this overview that the usual notion of rank and other intuitions from linear algebra are valid over $\mathbb{Z}_m$ and let us proceed with sketching the proof of Theorem 2 using the equivalent formulation as bounding (from be-

low) the rank of a MV matrix P. The starting point is a generalization of a result of [20], mentioned above, from $\mathbb{Z}_2$ to $\mathbb{Z}_m$. We show that every matrix P over $\mathbb{Z}_m$ that is biased (i.e., its values are not distributed close to uniformly) and has low rank, contains a large monochromatic sub-matrix *modulo some factor* m' *of* m. The size of the sub-matrix is bounded from below by $\sim |P| \exp(-r'/\log(r'))$, where r' is the rank of P modulo m' (this factor depends on the specific way the matrix is biased). This generalizes the result of [20] which assumes $m = 2$ and finds a large monochromatic sub-matrix (modulo 2). We note that the sub-matrix lemma is the only component in the proof that relies on the PFR conjecture. Let us refer to this result from now on as the *sub-matrix* lemma. We can apply the sub-matrix lemma to a MV matrix P since its values are far from uniform (the probability of zero is much less than $1/m$) and since its rank is assumed (towards a contradiction) to be low.

Suppose for the sake of simplicity that $m = p \cdot q$, with p, q distinct primes (the proof for general m is significantly more technical but relies on the same basic intuitions). Applying the sub-matrix lemma we obtain a sub-matrix P_1 of P that is constant modulo some factor m_1 of m (so m_1 is either p, q or m) of size at least $|P| \exp(-r_1/\log(r_1))$, where $r_1 \leq n$ is the rank of $P \mod m_1$. Using some matrix manipulations, and subtracting a rank one matrix, we can get a large sub-matrix P_1' that does not intersect the diagonal of P and s.t all of the entries of P_1' are zero modulo m_1. Suppose $|P_1'| = t_1$ and consider the $2t_1 \times 2t_1$ sub-matrix P_1'' of P that has P_1' as its top-right (or bottom-left) block and s.t the top-left and bottom-right blocks are taken to have zero diagonal elements. Formally, if P_1' is indexed by rows

in R and columns in T with $R \cap T = \emptyset$ then the rows/columns of P_1'' will be indexed by $R \cup T$. If we consider the matrix P_1'' modulo m_1 then it has top-right block which is all zero and so its rank (modulo m_1) will be the sum of the ranks of the top-left and bottom right blocks. Thus, one of these blocks, w.l.o.g the top-left one, must have rank at most $n/2$ (over $\mathbb{Z}_{m_1}$). Notice also that both of these blocks are themselves MV matrices modulo m since they are sub-matrices of P with the same row and column sets. Let $\tilde{P}_1$ be the top-left block of P_1''. We can now apply, again, the monochromatic sub-matrix lemma to find a large sub-matrix P_2 of $\tilde{P}_1$ which is constant modulo some other factor m_2 of m. The size of P_2 will be

$$t_1 \cdot \exp(-r_2/\log(r_2)) = |P| \cdot \exp(-r_1/\log(r_1) - r_2/\log(r_2)).$$

The factor m_2 is also either p or q. If it happens to be that $m_1 = m_2$ then $r_2 \leq n/2$ and so we gain in the size of P_2 in this second step (the expression $r_2/\log(r_2)$ is smaller than $n/2\log(n/2)$ which is smaller by roughly a factor of two than our bound on $r_1/\log(r_1)$. Suppose we continue with this iterative process of finding constant sub-matrices for ℓ steps and that, by luck, all the factors $m_1, m_2, \ldots$ are equal to the same factor of m (say p). Then, after roughly $\log(n)$ iteration, we will reduce the rank modulo p to one and still have at least

$$|P| \cdot \exp\left(-\sum_{i=1}^{\ell} \frac{n}{2^i \log(n/2^i)}\right)$$

rows, which is close to the original size of P if we assume (in contradiction) that $|P| >> \exp(n/\log n)$. In this case we obtain a new large MV family U', V' modulo m such that all inner products $\langle u_i' \rangle v_j'$ of elements $u_i' \in U', v_j' \in V'$ are fixed modulo

p. From this we can easily construct a MV family of roughly the same size in $\mathbb{Z}_q^n$ and then use the bounds on $\mathbf{MV}(q, n)$ for primes to get a contradiction. In the 'unlucky' case we will have different factors $m_1, m_2, \ldots$ in each stage, but we can adapt the analysis to consider the decrease in rank simultaneously for all factors of m.

The full proof is by induction on the number of factors of m and uses the iterative sub-matrix argument above to go from a MV family modulo m to a MV family of roughly the same size modulo some proper factor of m (and then uses the inductive hypothesis on this new MV family).

2.1.3 Matrix rank over $\mathbb{Z}_m$

An important technical issue, which was already hinted at above, is in the definition of the rank of a matrix with entries in a ring $\mathbb{Z}_m$. There are two main properties of matrix rank over a field that we relied on in the proof sketch above. The first is that a rank r matrix is always the inner product matrix of vectors in r dimensions. Equivalently, a $t \times t$ matrix of rank r can be written as a product of a $t \times r$ matrix and an $r \times t$ matrix. This is important if we are to go back and forth between matrices and MV families. Another property we used is that, if we have a $2t \times 2t$ matrix composed of 4 blocks of size $t \times t$ and the top-right block is zero, then the rank of the matrix is the sum of the ranks of the top-left block and the bottom right block.

Ideally, we would like to define rank over $\mathbb{Z}_m$ so that both properties are satisfied. This is, however, impossible as the following example shows: Consider

the 2×2 matrix with the two rows $(4, 0)$ and $(0, 3)$ over $\mathbb{Z}_6$. This matrix can be written as the product of the two vectors $(2, 3)^T$ and $(2, 3)$ and so should have rank one, if we are to satisfy the first property. However, if we are to satisfy the second property, its rank should be the sum of the ranks of the two 1×1 matrices (4) and (3), which clearly cannot have rank zero!

Our solution to this problem is to give two different definitions of rank, each satisfying one of the two properties. We then show that the two definitions of rank can differ from each other by a multiplicative factor of $\log m$, which our proof can handle. The first definition of rank is as the smallest r such that our $t \times t$ matrix can be written as a product of a $t \times r$ matrix and an $r \times t$ matrix. Clearly this would satisfy the first property (but not the second). The second definition of rank is termed *column-rank* and is defined as the logarithm to the base m of the size of the additive subgroup of $\mathbb{Z}_m^t$ generated by the columns of the matrix. Notice that this definition of rank can result in the rank being non-integer. For example, the rank of the matrix with a single column $(2, 0)$ over $\mathbb{Z}_6$ would be equal to $\log_6(3)$ since the subgroup generated by this column is composed of the three vectors $(2, 0), (4, 0), (0, 0)$. It is not hard to show (see Claim 2.4.9) that this definition satisfies the second property described above regarding block matrices. Clearly, the two definitions agree for matrices over a field. We show (see Claim 2.4.6) that the two notions of rank can differ by a multiplicative factor of at most $\log m$. This allows us to use both definitions in different parts of the proof without losing too much in the transition. We finish this discussion by noting that in no part of the proof do we use the characterization of rank using determinants, which is often very

useful when working over a field.

2.1.4 Subsequent work

If q is a constant, Theorem 1 is of the form $\mathbf{MV}(m, n, q) \leq cm^{n/2}$ for some constant c. However, for the case of unrestricted q, ($q \leq m$), Theorem 1 gives an upper bound of $\mathbf{MV}(m, n) = m^{n/2+O(\log m)}$. Subsequent to our work, this was improved to an $O(m^{n/2+8.47})$ bound in [57].

2.1.5 Organization

We begin with some preliminaries in Section 2.2. We prove Theorem 1 in Section 2.3. Section 2.4 contains some claims about matrices over $\mathbb{Z}_m$. Section 2.5 introduces collision free MV families. Both Section 2.4 and Section 2.5 will be used in the proof of Theorem 2 in Section 2.6. The proof of Theorem 2 also requires the sub-matrix lemma, whose proof appears in Section 2.7.

2.2 General preliminaries

Notations: Throughout the paper we will be handling ordered lists of elements. A list A of size t over a finite set Ω is an ordered t-tuple $A = (a_1, a_2, \cdots, a_t)$ where each $a_i \in \Omega$. A list can have repetitions. If it doesn't we say it is *twin free*. When discussing sublists $A \subseteq B$ with $B = (b_1, \ldots, b_t)$ we will use the convention that, unless specified otherwise, A maintains the ordering induced by B. For a positive integer t, we let $[t]$ denote the list $(1, \cdots t)$. So, for example, when we say that $T \subseteq [t]$ we mean that T is a list of integers in increasing order belonging to $[t]$. We say that a list $A = (a_1, \ldots, a_t)$ over Ω is *constant* if $a_i = a_j$ for all $i, j \in [t]$. We assume all logarithms are in base 2 unless otherwise specified.

2.2.1 MV Families: Basic Facts and Definitions

We now start with some basic definition and claims regarding MV families.

Definition 2.2.1 (Matching Vector Family)**.** *Let* $U = (u_1, u_2, \cdots u_t)$ *and* $V = (v_1, v_2, \cdots v_t)$ *be lists over* $\mathbb{Z}_m^n$*. Then* (U, V) *is called a* matching vector family *of size* t *in* $\mathbb{Z}_m^n$ *if*

- $\langle u_i, v_i \rangle = 0 \pmod m, \quad \forall i.$
- $\langle u_i, v_j \rangle \neq 0 \pmod m, \quad \forall i \neq j.$

If in addition, we $|\{\langle u, v \rangle : u \in U, v \in V\}| = q$*, we call such a MV family an* $q-$*restricted MV family. We denote the size of* (U, V) *by* $|(U, V)|$*. For instance,* $|(U, V)| = t$ *above.*

Definition 2.2.2 (Subset of Matching Vector Family)**.** *Let* $U = (u_1, u_2, \cdots u_t), V = (v_1, v_2, \cdots v_t)$ *form a matching vector family in* $\mathbb{Z}_m^n$ *of size* t*. By* $(U', V') \subseteq (U, V)$*, we mean there exists a sublist* $T \subseteq [t]$ *such that* $U' = (u_i : i \in T), V' = (v_i : i \in T)$*. Observe that* (U', V') *is a matching vector family in* $\mathbb{Z}_m^n$*.*

Definition 2.2.3 ($\mathbf{MV}(m, n)$)**.** *We denote by* $\mathbf{MV}(m, n)$ *the maximum size of a matching vector family* (U, V) *in* $\mathbb{Z}_m^n$*. Similarly, we denote by* $\mathbf{MV}(m, n, q)$ *the maximum size of an* $q-$*restricted matching vector family* (U, V) *in* $\mathbb{Z}_m^n$*.*

We shall use the following simple facts implicitly throughout the paper.

Fact 2.2.4. $\mathbf{MV}(m, n)$ *is an increasing function of* n*.*

Proof. For $n_1 < n_2$, we show $\mathbf{MV}(m, n_1) \leq \mathbf{MV}(m, n_2)$. Given (U, V), a matching vector family in $\mathbb{Z}_m^{n_1}$, we can pad each element in U and V by $n_2 - n_1$ zeros and obtain a matching vector family in $\mathbb{Z}_m^{n_2}$ of the same size. □

Fact 2.2.5. *If (U, V) is a matching vector family in $\mathbb{Z}_m^n$, then U and V are twin free.*

Proof. Let $U = (u_1, u_2, \cdots u_t), V = (v_1, v_2, \cdots v_t)$. We prove U is twin free. By symmetry V is also twin free. Suppose $u_i = u_j$ for some $i \neq j$. Now, $\langle u_i, v_j \rangle = \langle u_j, v_j \rangle = 0$ which is a contradiction. □

To facilitate writing in the proofs to follow we introduce the following notation for taking lists, matrices, etc. modulo an integer r.

Definition 2.2.6 (Modulo r notation)**.** *Let $2 \leq r \leq m$ be such that r divides m. Given $a = (a_1, \cdots, a_n) \in \mathbb{Z}_m^n$, we denote by $a^{(r)} = (a_1 \pmod r, \cdots, a_n \pmod r) \in \mathbb{Z}_r^n$. For a list $U = (u_1, u_2, \cdots u_t)$ over $\mathbb{Z}_m^n$, let $U^{(r)} = \left(u_1^{(r)}, u_2^{(r)}, \cdots u_t^{(r)}\right)$. Also, if $u^{(r)}$ is constant for all $u \in U$, we say $U^{(r)}$ is constant. Similarly, for a $t \times t$ matrix M over $\mathbb{Z}_m$, define $M^{(r)}$ to be the $t \times t$ matrix over $\mathbb{Z}_r$ such that $M^{(r)}(j, k) = M(j, k) \pmod r$ for all $1 \leq j, k \leq t$.*

We will also need the following definitions.

Definition 2.2.7 (Bucket $B_r(w, A)$)**.** *Let $A \subseteq \mathbb{Z}_m^n$ be a list. For any $w \in \mathbb{Z}_r^n$, we denote by $B_r(w, A) = \left(a \in A : a^{(r)} = w\right)$ the sub-list of elements of A which are equal to w modulo r.*

Definition 2.2.8 (Matrix $P_{U,V}$)**.** *Let $U = (u_1, u_2, \cdots u_t)$ and $V = (v_1, v_2, \cdots v_t)$ be lists over $\mathbb{Z}_m^n$. We let $P_{U,V}$ be the $t \times t$ matrix over $\mathbb{Z}_m$ defined by $P_{U,V}(i,j) = \langle u_i, v_j \rangle$ for $1 \leq i, j \leq t$.*

We will use the following lemma from [56] mentioned informally in the introduction.

Lemma 2.2.9. *[56, Theorem 21] For any positive integer n and prime p, $\mathbf{MV}(p,n) \leq 1 + \binom{n+p-2}{p-1}$.*

2.2.2 Probability Distributions

Definition 2.2.10. *For a distribution μ over a finite set Ω, we write $X \sim \mu$ to denote a random variable X drawn according to μ. We will also treat μ as a function $\mu : \Omega \mapsto [0,1]$ such that $\mu(x) = \mathbf{Pr}[X = x]$. For a list A over Ω, $x \sim A$ denotes a point sampled as per the uniform distribution on A (taking repetitions into account).*

Definition 2.2.11 (Statistical distance between distributions)**.** *Let μ_1 and μ_2 be two distributions over a finite set Ω. The* statistical distance *(or simply distance) between μ_1 and μ_2, denoted $\Delta(\mu_1, \mu_2)$, is defined as*

$$\Delta(\mu_1, \mu_2) = \frac{1}{2} \sum_{x \in \Omega} |\mu_1(x) - \mu_2(x)| .$$

Definition 2.2.12 (Collision probability)**.** *Given a distribution μ over a finite set Ω the* collision probability *of μ, denoted* $\mathrm{cp}(\mu)$*, is defined as*

$$\mathrm{cp}(\mu) = \mathbf{Pr}_{x,y \sim \mu}[x = y] = \sum_{x \in \Omega} \mu(x)^2 .$$

The following two lemmas are standard and their proofs are skipped.

Lemma 2.2.13. *Let μ be a distribution over $\mathbb{Z}_m$ and let $\mathbb{U}_m$ denote the uniform distriution over $\mathbb{Z}_m$. If $\Delta(\mu, \mathbb{U}_m) \geq \epsilon$ then for some $1 \leq j \leq m-1$,*

$$\left|\mathbb{E}_{x\sim\mu}\left[\left(\omega^j\right)^x\right]\right| \geq \frac{2\epsilon}{\sqrt{m}},$$

where $\omega = exp(2\pi i/m)$ is a primitive root of unity of order m.

Lemma 2.2.14. *Let ω be a primitive root of unity of order m. Let μ_1 and μ_2 be two probability distributions over $\mathbb{Z}_m^n$. If $\left|\mathbb{E}_{x\sim\mu_1, y\sim\mu_2}\left[\omega^{\langle x,y\rangle}\right]\right| \geq \epsilon$, then $\mathrm{cp}(\mu_1)\,\mathrm{cp}(\mu_2) \geq \epsilon^2/m^n$.*

2.3 Proof of Theorem 1

In this section we prove Theorem 1, restated here with explicit constants.

Theorem 2.3.1. *Let $m \geq 2, 2 \leq q \leq m$ and n be arbitrary positive integers. Then*

$$\mathbf{MV}(m,n,q) \leq \min\{12q \cdot q^{24\left(1+\log q^{10q}\right)} m^{n/2}, 12q \cdot q^{24(\log m)} m^{n/2}\}$$

For the purpose of the proof, we introduce the following notation that will be used only in this section.

Definition 2.3.2 ($\mathbf{MV}_{r_1,r_2}(m,n,q)$)**.** *Let r_1, r_2 be integers such that $r_1 r_2 | m$. We denote by $\mathbf{MV}_{r_1,r_2}(m,n,q)$ the maximum size of a $q-$restricted MV family (U,V) in $\mathbb{Z}_m^n$ satisfying*

- *$U^{(r_1)}$ and $V^{(r_2)}$ are constants.*

- $\langle u, v\rangle = 0 \pmod{r_1 r_2}$ *for all* $u \in U, v \in V$.

Note that $\mathbf{MV}_{1,1}(m, n, q) = \mathbf{MV}(m, n, q)$ *(with the convention that* $x \pmod 1 = 0$ *for any integer* x*).*

Before we go to the proof of Theorem 2, we have the following claims.

Claim 2.3.3. *Let* (U, V) *be a* q*-restricted matching vector in* $\mathbb{Z}_m^n$*. Then, without loss of generality,* m *has at most* q *prime factors.*

Proof. Assume $m = \prod_{i=1}^{r} p_i^{e_i}$ with possible $r > q$. Let $v_1, \cdots, v_q \in Z_m$ be the q possible nonzero values that the inner products $\langle u, v\rangle$ attain. For each v_j there is some prime p_{i_j} where $v_j \neq 0 \left(\mod p_{i_j}^{e_{i_j}} \right)$. So, we can replace m with just $\prod_{j=1}^{q} p_{i_j}^{e_{i_j}}$ and discard all primes other than $p_{i_1}, \cdots, p_{i_q}$. □

Claim 2.3.4. *If* N *has* r *prime factors, then* $|\{x \in Z_N : order(x) < N/S\}| \leq N/S \cdot (\log S)^r$.

Proof. Assume $N = \prod_{i=1}^{r} p_i^{e_i}$. An element x with $order(x) \leq N/S$ is divisible by some $\prod_{i=1}^{r} p_i^{f_i} \geq S$. Let $T = \{(f_1, \cdots, f_r) : \prod p_i^{f_i} \geq S\}$. Define a partial order on T by $(f_1, \cdots, f_r) \leq (f_1', \cdots, f_r')$ if $f_i \leq f_i'$. Let T' be a subset of T such that for any $t \in T$ there is $t' \in T'$ such that $t' \leq t$. Note that if x has order $\leq N/S$ then x must be divisible by $\prod_i p_i^{f_i}$ for some $(f_1, \cdots, f_r)$ in T'. So, the number of elements of order $< N/S$ is at most $N|T'|/S$. We can bound the size of T' as follows: any element f_i is between 0 and $\log_{p_i} S$, since clearly if f_i is larger we can reduce f_i by one. So, $|T'| \leq \prod_{i=1}^{r} (\log S / \log p_i) <= (logS)^r$. □

The proof of Theorem 2.3.1 will follow immediately from the following two lemmas, which will be proved below.

Lemma 2.3.5. *Let $m = r_1 r_2 r_3$ where r_1, r_2, r_3 are arbitrary positive integers such that $r_3 \geq 2$. Let $q \geq 2, t \geq 12q$ and n be arbitrary positive integers. Let (U, V) be a $q-$restricted matching vector family in $\mathbb{Z}_m^n$ with $|(U, V)| = t$ such that*

- *$U^{(r_1)}$ and $V^{(r_2)}$ are constants.*
- *$\langle u, v \rangle = 0 \pmod{r_1 r_2}$ for all $u \in U, v \in V$.*

Then, there exists $s | r_3$ with $s \geq \max\{2, r_3/q^{10q}\}$ and a $q-$restricted matching vector family $(U', V') \subseteq (U, V)$ such that $|(U', V')| \geq s^{-n/2} q^{-24} t$ where

- *$\langle u', v' \rangle = 0 \pmod{r_1 r_2 s}$ for all $u' \in U', v' \in V'$.*
- *Either $U'^{(r_1 s)}$ is constant or $V'^{(r_2 s)}$ is constant.*

Applying Lemma 2.3.5 iteratively we can prove the following bound.

Lemma 2.3.6. $\mathbf{MV}_{r_1, r_2}(m, n, q) \leq 12q \cdot q^{24 \log \frac{m}{r_1 r_2}} \left(\frac{m}{r_1 r_2}\right)^{n/2}$.

Given Lemma 2.3.6 and Lemma 2.3.5, we now show how to deduce Theorem 2.3.1.

Proof of Theorem 2.3.1. Observe that for any matching vector family (U, V) in $\mathbb{Z}_m^n$, $U^{(1)}$ and $V^{(1)}$ are constants and $\langle u, v \rangle = 0 \pmod 1$ for all $u \in U, v \in V$. Thus, $\mathbf{MV}(m, n, q) = \mathbf{MV}_{1,1}(m, n, q)$. Using this observation and setting $r_1 = r_2 = 1$

in Lemma 2.3.6 gives us the second bound in the main theorem. We now proceed to prove the first bound. *Case 1:* $m \leq q^{10q}$. Applying Lemma 2.3.6, we get $\mathbf{MV}(m,n,q) = \mathbf{MV}_{1,1}(m,n,q) \leq 12q \cdot q^{24\log q^{10q}}(m)^{n/2} \leq 12q \cdot q^{24(1+\log q^{10q})}(m)^{n/2}$.

Case 2: $m > q^{10q}$. By Lemma 2.3.6, we know that for $s \geq m/q^{10q}$, $\mathbf{MV}_{1,s}(m,n,q) \leq 12q \cdot q^{24\log \frac{m}{s}}\left(\frac{m}{s}\right)^{n/2} \leq 12q \cdot q^{24\log q^{10q}}\left(\frac{m}{s}\right)^{n/2}$. Similarly, we have for $s \geq m/4q$, $\mathbf{MV}_{s,1}(m,n,q) \leq 12q \cdot q^{24\log q^{10q}}\left(\frac{m}{s}\right)^{n/2}$.

Now, suppose there is a $q-$*restricted* MV family (U,V) in $\mathbb{Z}_m^n$ of size $t > 12q \cdot q^{24\left(1+\log q^{10q}\right)}m^{n/2}$. Applying Lemma 2.3.5 with $r_1 = r_2 = 1$, we get a $q-$*restricted* MV family $(U',V') \subseteq (U,V)$ of size $t' \geq s^{-n/2}q^{-24}t > q^{24\log q^{10q}}\left(\frac{m}{s}\right)^{n/2}$ where $s \geq m/q^{10q}$ such that

- $\langle u', v'\rangle = 0 \ (mod\ s)$ for all $u' \in U', v' \in V'$.
- Either $U'^{(s)}$ is constant or $V'^{(s)}$ is constant.

But, by the previous paragraph, we have for $s \geq m/q^{10q}$, $\mathbf{MV}_{s,1}(m,n,q)$ and $\mathbf{MV}_{1,s}(m,n,q)$ are at most $12q \cdot q^{24\log q^{10q}}\left(\frac{m}{s}\right)^{n/2}$. This leads to a contadiction. □

2.3.1 Proof of Lemma 2.3.5

By assumption we have that $\langle u,v\rangle = 0 \ (mod\ r_1r_2)$ for all $u \in U, v \in V$. So, we can consider $\frac{\langle u,v\rangle}{r_1r_2} \in \mathbb{Z}_{r_3}$. Also, by hypothesis, the inner products $\frac{\langle u,v\rangle}{r_1r_2}$ occupy $q' \leq q$ residues in $\mathbb{Z}_{r_3}$. We have that

- For $1 \leq i \leq t$, $\frac{\langle u_i,v_i\rangle}{r_1r_2} = 0 \ (mod\ r_3)$ since $\langle u_i, v_i\rangle = 0 \pmod m$.
- For $1 \leq i,j \leq t$, $i \neq j$, $\frac{\langle u_i,v_j\rangle}{r_1r_2} \neq 0 \ (mod\ r_3)$ since $\langle u_i, v_j\rangle \neq 0 \pmod m$.

Let μ denote the distribution over $\mathbb{Z}_{r_3}$ defined by $\frac{\langle u_i, v_j \rangle}{r_1 r_2} \mod r_3$ where u_i, v_j are drawn independently and uniformly from U, V respectively.

Case 1: $4q' \geq r_3$. Observe that μ outputs 0 only when $i = j$. Therefore, $\mathbf{Pr}[\mu = 0] = 1/t \leq 1/12q' \leq 1/3r_3$. On the other hand, $\mathbf{Pr}[\mathbb{U}_{r_3} = 0] = 1/r_3$. This implies that $\Delta(\mu, \mathbb{U}_{r_3}) \geq 1/3r_3$. Thus, applying Lemma 2.2.13 with $\omega = exp(2\pi i/r_3)$, we get that for some $1 \leq j \leq r_3 - 1$,

$$\left|\mathbb{E}_{x\sim\mu}\left[\left(\omega^j\right)^x\right]\right| \geq \frac{2}{3r_3\sqrt{r_3}} \geq \frac{1}{12q'^{3/2}}.$$

Let $\omega' = \omega^j$ and $ord(\omega')$ (the order of ω') be $s = r_3/gcd(r_3, j)$. Also, note that as $j \geq 1$, we have $s \geq 2$. Also, trivially, $s \geq r_3/q'^{10q'} \geq r_3/q^{10q}$.

Case 2: $4q' < r_3$. Let X be the random variable that picks a random $0 \leq j \leq r_3 - 1$ and outputs $\left|\mathbb{E}_{x\sim\mu}\left[(\omega^j)^x\right]\right|$. We will now show that with significant probability $X^2 \geq 1/2q'$. First observe that $X \leq 1$. On the other hand, we will show that $E[X^2]$ is large. To see this, let $Z = \{z_1, \cdots z_{q'}\}$ be the q' residues forming the support of μ. Also, for $1 \leq i \leq q'$, let $\alpha_i \stackrel{def}{=} \mu(z_i)$. Then,

$$\begin{aligned}
\mathbb{E}_j\left[X^2\right] &= \mathbb{E}_j\left[\sum_{1\leq i,i'\leq q'} \alpha_i \alpha_{i'} \omega^{j(z_i - z_{i'})}\right] \\
&= \sum_{1\leq i\leq q'} \alpha_i^2 \\
&\geq 1/q'
\end{aligned}$$

Therefore, we claim that $\mathbf{Pr}[X^2 \geq 1/2q'] \geq 1/2q' \geq 1/2q$. If not, then

$$\begin{aligned}
E_j\left[X^2\right] &= \mathbf{Pr}[X^2 \geq 1/2q'] E_j\left[X^2 | X^2 \geq 1/2q'\right] + \mathbf{Pr}[X^2 < 1/2q'] E_j\left[X^2 | X^2 < 1/2q'\right] \\
&< 1/2q' + 1/2q' \\
&= 1/q'
\end{aligned}$$

which is a contradiction.

By the above, we already have that there exists some ω' such that $|\mathbb{E}_{x\sim\mu}[(\omega')^x]| \geq 1/\sqrt{2q'}$ and $ord(\omega') \geq 2$ since $r_3/2q' > 1$ and thus ω' is not trivial.

Now, we shall show the existence of ω' of much higher order provided $r_3 > q'^{10q'}$. By Claim 2.3.4, for $S = q'^{10q'}$ and $N = r_3$, and noting that r_3 has atmost q prime factors by Claim 2.3.3, we have

$$Pr_j[ord\left(\omega^j\right) \leq r_3/S] \leq 1/4q'$$

Thus, with probabilty at least $1/2q' - 1/4q' = 1/4q'$, a random j satisfies

- $\left|\mathbb{E}_{x\sim\mu}\left[(\omega^j)^x\right]\right| \geq 1/\sqrt{2q'} \geq \frac{1}{12q^{3/2}}$
- $s = ord(\omega^j) \geq r_3/S$

Also, as $r_3/4q' > 1$ the above two conditions are true for some $j \neq 0$.

Now, we combine the above two cases as follows. Let $\omega' = \omega^j$ and $\varepsilon = \frac{1}{12q^{3/2}}$. We have shown by the above case-by-case analysis that

- $|\mathbb{E}_{x\sim\mu}[(\omega')^x]| \geq \epsilon$
- $s = ord(\omega')$ is such that $s \geq \max\{2, r_3/q^{10q}\}$

Using the Cauchy-Schwartz inequality twice we get

$$\begin{aligned}
& \left|\mathbb{E}_{u\sim U,v\sim V}\left[\left(\omega'\right)^{\langle u,v\rangle/r_1r_2}\right]\right| \geq \epsilon \\
\implies & \left|\mathbb{E}_{u,\tilde{u}\sim U,v\sim V}\left[\left(\omega'\right)^{\langle u-\tilde{u},v\rangle/r_1r_2}\right]\right| \geq \epsilon^2 \\
\implies & \left|\mathbb{E}_{u,\tilde{u}\sim U,v,\tilde{v}\sim V}\left[\left(\omega'\right)^{\langle u-\tilde{u},v-\tilde{v}\rangle/r_1r_2}\right]\right| \geq \epsilon^4 \\
\implies & \left|\mathbb{E}_{u,\tilde{u}\sim U,v,\tilde{v}\sim V}\left[\left(\omega'\right)^{\langle (u-\tilde{u})/r_1,(v-\tilde{v})/r_2\rangle}\right]\right| \geq \epsilon^4.
\end{aligned}$$

We need to explain the last expression. Since by assumption $U^{(r_1)}$ and $V^{(r_2)}$ are constants, $(u-\tilde{u})/r_1 \in \mathbb{Z}_m^n$ and $(v-\tilde{v})/r_2 \in \mathbb{Z}_m^n$ are well defined. Thus, we can fix $\tilde{u}$ and $\tilde{v}$ by an averaging argument such that

$$\left|\mathbb{E}_{u\sim U,v\sim V}\left[\left(\omega'\right)^{\langle (u-\tilde{u})/r_1,(v-\tilde{v})/r_2\rangle}\right]\right| \geq \epsilon^4.$$

Let $U' = (u'_1, u'_2, \cdots u'_t), V' = (v'_1, v'_2, \cdots v'_t)$ where $u'_i = (u_i - \tilde{u})/r_1$ and $v'_i = (v_i - \tilde{v})/r_2$. Notice that U' and V' are not assumed to be a MV family (later we will derive from them a MV family). We now define two probability distributions $\mu^{U'}$ and $\mu^{V'}$ over $\mathbb{Z}_s^n$. For each $w \in \mathbb{Z}_s^n$, let $\mu^{U'}(w) = |B_s(w, U')| / |U'|$ and $\mu^{V'}(w) = |B_s(w, V')| / |V'|$. That is, $\mu^{U'}(w)$ is the probability that $u'^{(s)} = w$ where u' is chosen uniformly in U', and similarly for $\mu^{V'}(w)$. Therefore, since the order of w' is s, we have that

$$\left|\mathbb{E}_{w_1\sim\mu^{U'},w_2\sim\mu^{V'}}\left[\left(\omega'\right)^{\langle w_1,w_2\rangle}\right]\right| \geq \epsilon^4.$$

Recalling that s is the order of ω' and applying Lemma 2.2.14, we get $\mathrm{cp}\left(\mu^{U'}\right)\mathrm{cp}\left(\mu^{V'}\right) \geq \epsilon^8/s^n$. Therefore, one of $\mathrm{cp}\left(\mu^{U'}\right)$, $\mathrm{cp}\left(\mu^{V'}\right)$, say $\mathrm{cp}\left(\mu^{U'}\right)$, is

at least $\epsilon^4/s^{n/2}$. Let w^* be the point of maximum probability mass given by $\mu^{U'}$. Then,

$$\mu^{U'}(w^*) = \mu^{U'}(w^*) \sum_{w \in \mathbb{Z}_s^n} \mu^{U'}(w) \geq \sum_{w \in \mathbb{Z}_s^n} \mu^{U'}(w)^2 = \mathrm{cp}\left(\mu^{U'}\right) \geq \epsilon^4/s^{n/2}.$$

Now, $\mu^{U'}(w^*) \geq \epsilon^4/s^{n/2}$ means that $\left|\{u \in U : \frac{u-\tilde{u}}{r_1} = w^* \pmod{s}\}\right| \geq t\epsilon^4/s^{n/2}$. Equivalently,

$$\left|\{u \in U : u - \tilde{u} = r_1 w^* \pmod{r_1 s}\}\right| \geq t\epsilon^4/s^{n/2}.$$

Let $T' = (i : u_i = \tilde{u} + r_1 w^* \pmod{r_1 s})$. Now, define $U'' = (u_i : i \in T')$ and $V'' = (v_i : i \in T')$. Observe that (U'', V'') is a matching vector family in $\mathbb{Z}_m^n$ such that

- $U''^{(r_1 s)}$ and $V''^{(r_2)}$ are constants.
- $|(U'', V'')| \geq t\left(\epsilon^4/s^{n/2}\right)$.

The only thing left is to show that $\langle u, v \rangle = 0 \pmod{r_1 r_2 s}$ for all $u \in U'', v \in V''$. This may not be true in general. However, we can take a large subset of the matching vector family so that the resulting matching vector family satisfies this condition. To see this, let $u \in U'', v \in V''$ be arbitrary. Now, $u = r_1 s \cdot u' + u_0$ and $v = r_2 \cdot v' + v_0$ where u', v' depend on u, v respectively and u_0, v_0 are independent of u, v. Then,

$$\langle u, v \rangle = r_1 r_2 s \langle u', v' \rangle + r_1 s \langle u', v_0 \rangle + r_2 \langle u_0, v' \rangle + \langle u_0, v_0 \rangle.$$

As u varies over U'', $\langle u', v_0 \rangle$ takes at most q values modulo r_2. Hence, $r_1 s \langle u', v_0 \rangle$ takes at most q values modulo $r_1 r_2 s$. Therefore, there exist at least $(1/q)|U''|$ elements

of U'' such that $r_1 s\langle u', v_0\rangle$ is a constant modulo $r_1 r_2 s$. We take the corresponding elements from V'' to form a matching vector family $(U''', V''') \subseteq (U'', V'')$. We apply another round using the same idea on U''', V''', this time ensuring that $r_2\langle u_0, v'\rangle$ is constant modulo $r_1 r_2 s$ as v varies over a large fraction of V'''. Thus, we end up with $\tilde{V}$ of size at least $(1/q)\,|V'''|$ such that $r_2\langle u_0, v_i\rangle$ is a constant modulo $r_1 r_2 s$. We take the corresponding subset $\tilde{U}$ from U''' so that $(\tilde{U}, \tilde{V}) \subseteq (U''', V''')$ is a matching vector family. Denote the size of $(\tilde{U}, \tilde{V})$ by $\tilde{t}$. Note that $\tilde{U} = (\tilde{u}_1, \cdots, \tilde{u}_{\tilde{t}})\,, \tilde{V} = (\tilde{v}_1, \cdots, \tilde{v}_{\tilde{t}})$ is a matching vector family in $\mathbb{Z}_m^n$ of size at least $(1/q^2)\,t\left(\epsilon^4/s^{n/2}\right) = s^{-n/2} q^{-\left(8+4\log_q(12)\right)} t \geq s^{-n/2} q^{-(8+4\log_2(12))} t \geq s^{-n/2} q^{-24} t$. Also, as $\langle u, v\rangle$ is a constant modulo $r_1 r_2 s$, for $u \in \tilde{U}, v \in \tilde{V}$, and $\langle \tilde{u}_i, \tilde{v}_i\rangle = 0 \pmod{r_1 r_2 s}$, we get that $\langle u, v\rangle = 0 \pmod{r_1 r_2 s}$, for $u \in \tilde{U}, v \in \tilde{V}$. This concludes the proof. □

2.3.2 Proof of Lemma 2.3.6

We prove the lemma by backward induction on $r_1 r_2 | m$. That is, to prove the claim about $\mathbf{MV}_{r_1,r_2}(m, n, q)$, we assume the inductive hypothesis for $\mathbf{MV}_{r_1',r_2'}(m, n, q)$ where $r_1' r_2' > r_1 r_2$ and $r_1' r_2' | m$.

Base Case. The base case of $r_1 r_2 = m$ is trivial. To see this, observe that if $\langle u, v\rangle = 0 \pmod{m}$ for all $u \in U, v \in V$, then by the definition of a matching vector family in $\mathbb{Z}_m^n$, the size of such a family cannot exceed 1. Hence, for $r_1 r_2 = m$, $\mathbf{MV}_{r_1,r_2}(m, n, q) = 1 \leq 12q \cdot q^{24\log\frac{m}{r_1 r_2}} \left(\frac{m}{r_1 r_2}\right)^{n/2}$.

Inductive Step. Let $m = r_1 r_2 r_3$ with $r_1 r_2 < m$ (that is, $r_3 \geq 2$). By the inductive hypothesis we have $\mathbf{MV}_{r_1',r_2'}(m, n, q) \leq 12q \cdot q^{24\log\frac{m}{r_1' r_2'}} \left(\frac{m}{r_1' r_2'}\right)^{n/2}$ for all r_1', r_2' such that $r_1' r_2' > r_1 r_2$ and $r_1' r_2' | m$. We need to show that $\mathbf{MV}_{r_1,r_2}(m, n, q) \leq$

$12q \cdot q^{24\log\frac{m}{r_1r_2}}\left(\frac{m}{r_1r_2}\right)^{n/2}$. Suppose this is false, so that there exists a $q-restricted$ matching vector family (U,V) in $\mathbb{Z}_m^n$ with $U=(u_1,\cdots u_t), V=(v_1,\cdots v_t)$ where $t > 12q \cdot q^{24\log\frac{m}{r_1r_2}}\left(\frac{m}{r_1r_2}\right)^{n/2}$ such that

- $U^{(r_1)}$ and $V^{(r_2)}$ are constants.
- $\langle u,v\rangle = 0 \ (mod\ r_1r_2)$ for all $u \in U, v \in V$.

Note that $t \geq 12q$. Therefore, applying Lemma 2.3.5, there exists $s|r_3$ with $s \geq 2$ and matching vector family $(U',V') \subseteq (U,V)$ such that $|(U',V')| \geq s^{-n/2}q^{-24}t$ where

- $\langle u',v'\rangle = 0 \ (mod\ r_1r_2s)$ for all $u' \in U', v' \in V'$.
- either $U'^{(r_1s)}$ is constant or $V'^{(r_2s)}$ is constant.

Without loss of generality, we assume that $U'^{(r_1s)}$ is a constant. Therefore,

$$\begin{aligned}|(U',V')| &> s^{-n/2}q^{-24}\cdot 12q\cdot q^{24\log\frac{m}{r_1r_2}}\left(\frac{m}{r_1r_2}\right)^{n/2}\\ &= 12q\cdot q^{24\left(\log\frac{m}{r_1r_2}-1\right)}\left(\frac{m}{r_1r_2s}\right)^{n/2}\\ &\geq 12q\cdot q^{24\log\frac{m}{r_1r_2s}}\left(\frac{m}{r_1r_2s}\right)^{n/2},\end{aligned}$$

where the last inequality used the fact that $s \geq 2$. This however contradicts the inductive hypothesis. □

2.4 Matrices over $\mathbb{Z}_m$

Notations: For a $t \times s$ matrix M over $\mathbb{Z}_m$ and for lists $T \subseteq [t], S \subseteq [s]$ the $T \times S$ submatrix of M is the matrix with rows in T and columns in S. For $i \in [s]$ and $j \in [t]$ we denote the i'th row of M by $M(i:)$ and the j'th column by $M(:j)$.

Definition 2.4.1 (Span of a set)**.** *For $A \subseteq \mathbb{Z}_m^n$ let* $\operatorname{span}(A)$ *denote the additive subgroup generated by A. We say that a set A spans $u \in \mathbb{Z}_m^n$ if $u \in \operatorname{span}(A)$.*

Definition 2.4.2 (Rank of a matrix over $\mathbb{Z}_m$)**.** *Let M be a $t \times t$ matrix over $\mathbb{Z}_m$. Then* $\operatorname{rank}(M)$ *is the smallest r such that $M = AB$ where A is an $t \times r$ martrix over $\mathbb{Z}_m$ and B is an $r \times t$ matrix over $\mathbb{Z}_m$.*

Definition 2.4.3 (Column rank of a matrix over $\mathbb{Z}_m$)**.** *Let M be a $t \times t$ matrix over $\mathbb{Z}_m$. Let* $\operatorname{colspan}(M)$ *denote the subgroup of $\mathbb{Z}_m^t$ generated by the columns of M. The column rank of M over $\mathbb{Z}_m$ is defined as*

$$\operatorname{colrank}(M) = \log_m |\operatorname{colspan}(M)|.$$

The column rank is, in general, a real number in the range $[0, t]$.

Since the rank can behave in unexpected ways over $\mathbb{Z}_m$, we make sure to prove some of the basic facts that we will be using later on.

Fact 2.4.4. *Let M be a $t \times t$ matrix over $\mathbb{Z}_m$ and let M' be any submatrix of M. Then* $\operatorname{colrank}(M') \leq \operatorname{colrank}(M)$.

Proof. Suppose M' is given by the first t' rows and the first t'' columns of M. We will define an injective map $f : \operatorname{colspan}(M') \to \operatorname{colspan}(M)$. Given any $x \in \operatorname{colspan}(M')$ we can write $x = \sum_{j=1}^{t''} \alpha_j \cdot M'(:\, j)$ in some fixed way (there might be several choices of α_j). Define $f(x) = \sum_{j=1}^{t''} \alpha_j \cdot M(:\, j)$. Then, x is clearly the restriction of $f(x)$ to the first t' indices and so the map is injective. □

Fact 2.4.5. *Let M be a $t \times t$ matrix over $\mathbb{Z}_m$ and let $s|m$. Then* $\operatorname{rank}\left(M^{(s)}\right) \leq \operatorname{rank}(M)$.

Proof. Suppose there exist an $t \times r$ matrix A and an $r \times t$ matrix B over $\mathbb{Z}_m$ such that $M = AB$. Then $M^{(s)} = A^{(s)}B^{(s)}$ and so the rank of $M^{(s)}$ is at most r. □

We will need the following claims relating the rank and the column rank of matrices over $\mathbb{Z}_m$.

Claim 2.4.6. *Let M be an $t \times t$ matrix over $\mathbb{Z}_m$. Then,*

$$\frac{\operatorname{rank}(M)}{\log m} \leq \operatorname{colrank}(M) \leq \operatorname{rank}(M).$$

Proof. Let $r = \operatorname{rank}(M)$ and $r' = \operatorname{colrank}(M)$. We first prove that $r' \leq r$. This is equivalent to proving that $|\operatorname{colspan}(M)| \leq m^r$. Let $M = AB$ where A is an $t \times r$ martrix over $\mathbb{Z}_m$ and B is an $r \times t$ matrix over $\mathbb{Z}_m$. Since the columns of M are all in the span of the columns of A we have that the column span of M can contain at most m^r elements.

We now prove that $r' \geq r/(\log m)$ or, equivalently, $|\operatorname{colspan}(M)| \geq 2^r$. Suppose in contradiction that $|\operatorname{colspan}(M)| < 2^r$. Take a minimal spanning set S of $\operatorname{colspan}(M)$ (that is, a set that spans $\operatorname{colspan}(M)$ and such that no proper subset of it does). Suppose $|S| \geq r$ and consider all linear combinations (over $\mathbb{Z}_m$) of elements of S with coefficients in $\{0, 1\} \subseteq \mathbb{Z}_m$. Since $|\operatorname{colspan}(M)| < 2^r$ there are two distinct $0-1$ linear combinations that map to the same element. This means that there is a linear combination with coefficients in $\{1, -1\}$ of the elements of S

that is equal to zero. Since both 1 and -1 are invertible modulo m we can write one of the elements of S as a linear combination of the other elements. This contradicts the minimality of S and so, we must have $|S| < r$. This implies that $\text{rank}(M) < r$, a contradiction, since we can write M as the product of the matrix with columns in S with the matrix of coefficients giving the columns of M. □

Claim 2.4.7. *Let M be an $t \times t$ matrix over $\mathbb{Z}_m$, let $r = \text{rank}(M)$. There exists r' columns of M that span the rest of M's columns such that $r' \leq r \log m$.*

Proof. Take a minimal spanning set S of the columns of M (that is, a set that spans all other columns and such that no proper subset of it spans all columns). If $2^{|S|} > m^r$, then $2^{|S|} > |\text{colspan}(M)|$ (by Claim 2.4.6) and we proceed as in the proof from Claim 2.4.6 above. If we look at all the $0-1$ combinations of the columns of S, then there are two distinct $0-1$ linear combinations of the columns that map to the same element of $\text{colspan}(M)$. Thus, let $\sum_i \alpha_i S(:i) = \sum_i \beta_i S(:i)$ where $\alpha_i \neq \beta_i$ for at least one i, say i_0. Therefore, we have $\sum_i (\alpha_i - \beta_i) S(:i) = 0$. Note that $(\alpha_{i_0} - \beta_{i_0}) = \pm 1$ and hence is invertible. This lets us write $S(:i_0)$ as a linear combinations of the remaining columns contradicting the minimality of S. Thus, $r' = |S| \leq r \log m$. □

The following claim shows that the column rank behaves similar to rank in terms of subadditivity.

Claim 2.4.8. *Let A, B be $t \times t$ matrices over $\mathbb{Z}_m$. Then,* $\text{colrank}(A+B) \leq \text{colrank}(A) + \text{colrank}(B)$.

Proof. We show that $|\text{colspan}(A+B)| \leq |\text{colspan}(A)|\,|\text{colspan}(B)|$. Note that $\text{colspan}(A+B) \subseteq \text{colspan}(A) + \text{colspan}(B) \stackrel{def}{=} \{a+b | a \in \text{colspan}(A), b \in \text{colspan}(B)\}$. Therefore, $|\text{colspan}(A+B)| \leq |\text{colspan}(A) + \text{colspan}(B)| \leq |\text{colspan}(A)|\,|\text{colspan}(B)|$. □

Claim 2.4.9. *Let M be a $2t \times 2t$ matrix over $\mathbb{Z}_m$, such that*

$$M = \begin{pmatrix} A & 0 \\ \star & B \end{pmatrix}$$

where A, B and $\star$ are $t \times t$ matrices. Then, $\text{colrank}(A) + \text{colrank}(B) \leq \text{colrank}(M)$.

Proof. We show that $|\text{colspan}(A)|\,|\text{colspan}(B)| \leq |\text{colspan}(M)|$. Let $\text{colspan}(A) = R_1$, $\text{colspan}(B) = R_2$, $\text{colspan}(M) = R$. We define $f : R_1 \times R_2 \to R$ and show that f is injective. Given $r_1 \in R_1$ and $r_2 \in R_2$, let $\alpha_1, \cdots \alpha_t$ and $\beta_1, \cdots \beta_t$ denote coefficients for linear combinations of the columns of A and B respectively that give r_1 and r_2. There might be many such linear combinations but we fix one for each r_i. Then, $f(r_1, r_2) = \sum_{i=1}^{t} \alpha_i M(:i) + \sum_{i=t+1}^{2t} \beta_{i-t} M(:i)$. Now, given a column vector $f(r_1, r_2) \in R$, we uniquely identify r_1 and r_2 as follows. We look at the first t rows and call it s_1. Now $s_1 = r_1$ and let $\alpha_1, \cdots \alpha_t$ be the linear combination fixed for r_1 while defining f. Now, consider $f(r_1, r_2) - \sum_{i=1}^{t} \alpha_i M(:i)$ and call the last t rows s_2. Note that $s_2 = r_2$. □

Claim 2.4.10. *Let M be a $t \times t$ square matrix over $\mathbb{Z}_m$ with zero diagonal entries. If for some $s|m$,* $\text{colrank}\left(M^{(s)}\right) \leq 2$, *then there exists at least $t' = t/m^2$ indices such that M restricted to those indices as rows and columns is the all zero matrix modulo s.*

Proof. As colrank $\left(M^{(s)}\right) \leq 2$, it follows that $\left|\text{colspan}\left(M^{(s)}\right)\right| \leq s^2 \leq m^2$. Hence, $M^{(s)}$ has at most m^2 distinct columns. Therefore, there exists a set of indices S of size $t' \geq t/m^2$ with $S = \{r_1, r_2, \cdots r_{t'}\}$ such that all the columns $M^{(s)}(: r_i)$ are identical. Also, as the diagonal elements are zero modulo m, they are zero modulo s. Thus, the $S \times S$ submatrix is the all zero matrix modulo s. □

2.5 Collision-Free MV families

In the proof of Theorem 2 it will be useful to assume that the elements of the MV family do not 'collide' when reduced modulo an integer s dividing m. In this section we develop the necessary machinery to allow for this assumption. We start by defining a collision free matching vector family.

Definition 2.5.1 (Collision free MV family)**.** *A* collision free matching vector family (U, V) *in* $\mathbb{Z}_m^n$ *is a matching vector family such that for all* $s|m, s \geq 2$*, all elements of* U *are distinct modulo* s*, and all elements of* V *are distinct modulo* s*. Note that if* (U, V) *is a collision free matching vector family, then so is any* $(U', V') \subseteq (U, V)$*.*

Lemma 2.5.2. *Let* $m \geq 2$ *be an arbitrary integer. Let* s *be a divisor of* m*, such that* $1 < s < m$*. Let* (U, V) *be a matching vector family in* $\mathbb{Z}_m^n$ *such that* $\langle u, v \rangle = 0 \pmod{s}$ *for all* $u \in U, v \in V$*. Then,* $|(U, V)| \leq \mathbf{MV}(m/s, n \log m)$*.*

Proof. Let $U = (u_1, u_2, \cdots u_t)$ and $V = (v_1, v_2, \cdots v_t)$. Recall that $P_{U,V}$ is the inner product matrix. We shall write $P_{U,V}$ as P in the rest of the proof for brevity. Let $r = \text{rank}(P) \leq n$. Hence, by Claim 2.4.7, there exists $r' \leq r \cdot \log m$ columns of P which span all the columns of P. As each entry of P is a multiple of s we can define a matrix P' over $\mathbb{Z}_{m/s}$ by $P' = (1/s) P$. We have

- $P'_{i,i} = 0 \quad \forall i.$
- $P'_{i,j} \neq 0 \quad \forall i \neq j.$

We next show that the r' columns that span the columns of P also span the columns in P'. Without loss of generality, let the first r' columns of P span the remaining columns of P. For any column j, let $P(:j) = \sum_{i=1}^{r'} c_i P(:i) \pmod m$. Since all entries of P are divisible by s, we can divide the expression by s and obtain that $P'(:j) = \sum_{i=1}^{r'} c_i P'(:i) \pmod{m/s}$. Hence, we deduce that $r_{P'} = \text{rank}(P') \leq r' \leq r \log m \leq n \log m$. This implies that $P' = AB$ for some $t \times r_{P'}$ matrix A and some $r_{P'} \times t$ matrix B over $\mathbb{Z}_{m/s}$. Thus, the rows of A and the columns of B form a matching vector family in $\mathbb{Z}_{m/s}^{r_{P'}}$. Therefore, $t \leq \mathbf{MV}(m/s, n \log m)$ as claimed. □

Lemma 2.5.3 (Bucket Lemma)**.** *For any m, let (U, V) be a matching vector family in $\mathbb{Z}_m^n$. Let $1 < s < m$ be any divisor of m. Then, for any $w \in \mathbb{Z}_s^n$, $|B_s(w, U)| \leq \mathbf{MV}(m/s, n \log m)$. By symmetry, $|B_s(w, V)| \leq \mathbf{MV}(m/s, n \log m)$.*

Proof. We prove that $|B_s(w, U)| \leq \mathbf{MV}(m/s, n)$. For $U = (u_1, u_2, \cdots u_t)$, consider any bucket $B_s(w, U) = U'$ *(say)*. Let $U' = \left(u_{j_1}, u_{j_2}, \cdots u_{j_{t'}}\right)$ where $1 \leq j_1 < j_2 < \cdots j_{t'} \leq t$. Let $V' = \left(v_{j_1}, v_{j_2}, \cdots v_{j_{t'}}\right)$. Now, for any $l, m \in [t']$, $\langle u_{j_l}, v_{j_l} \rangle = 0 \,(mod\ m)$. Therefore, $\langle u_{j_m}, v_{j_l} \rangle = 0 \,(mod\ s)$. By Lemma 2.5.2 on (U', V'), $t' \leq \mathbf{MV}(m/s, n \log m)$. □

We use the above lemma repeatedly to obtain a collision free matching vector family.

Lemma 2.5.4. *Let $m \geq 2$ be any positive integer. Suppose there is a matching vector family (U, V) in $\mathbb{Z}_m^n$. Then, there exists a collision free matching vector family $(U', V') \subseteq (U, V)$ such that*

$$|(U', V')| \geq \frac{|(U, V)|}{\left(\prod_{s|m, 1<s<m} \mathbf{MV}(s, n \log m)\right)^2}.$$

Proof. We will get rid of collisions iteratively by repeatedly applying Lemma 2.5.3. Let us write the divisors of m in ascending order as $2 \leq s_1 < s_2 < \cdots < s_l \leq m/2$. Perform the following operation for each $s|m$ starting from the smallest divisor greater than 1. For $0 \leq i \leq l$, let U_i, V_i be the matching vector after stage i with $U_0 = U$ and $V_0 = V$. Now suppose that we have U_i, V_i after the i'th stage such that there is no collision modulo s_j in U_i for $1 \leq j \leq i$. The $(i+1)$'th stage is performed as follows. Let us construct U_{i+1}, V_{i+1} from U_i, V_i to ensure no collision among the elements of U_{i+1} modulo s_{i+1} as well. For each $w \in \mathbb{Z}_{s_{i+1}}^n$, by Lemma 2.5.3, $\left|B_{s_{i+1}}(w, U_i)\right| \leq \mathbf{MV}(m/s_{i+1}, n \log m)$. Pick one element from each bucket in U_i and the corresponding matching vector from V_i to form (U_{i+1}, V_{i+1}). Thus, $|(U_{i+1}, V_{i+1})| \geq |U_i| / \mathbf{MV}(m/s_{i+1}, n \log m)$. We end up with matching vector family U_l, V_l such that $|(U_l, V_l)| \geq \frac{|(U,V)|}{\prod_{s|m, 1<s<m} \mathbf{MV}(m/s, n \log m)}$ and U_l is collision free. We repeat the same process this time pruning V_l in order to make it collision free as well. Thus, eventually we end up with a collision free matching vector family $(U_l', V_l') \subseteq (U, V)$ such that

$$|(U_l', V_l')| \geq \frac{|(U, V)|}{\left(\prod_{s|m, 1<s<m} \mathbf{MV}(m/s, n \log m)\right)^2} = \frac{|(U, V)|}{\left(\prod_{s|m, 1<s<m} \mathbf{MV}(s, n \log m)\right)^2}.$$

□

2.6 Proof of Theorem 2

Before proceeding with the proof we give yet another definition.

Definition 2.6.1. *Let $A, B \subseteq \mathbb{Z}_m^n$ be twin-free lists (or sets). Let ω be a primitive root of unity of order m. The duality measure of A, B with respect to ω is defined as*

$$D_\omega(A,B) = \left|\mathbb{E}_{a\sim A, b\sim B}\left[\omega^{\langle a,b\rangle}\right]\right|.$$

Notice that, if $\omega \neq 1$, $D_\omega(A,B) = 1$ implies that there is some $c \in \mathbb{Z}_m$ such that all the entries of the inner product matrix $P_{A,B}$ equal c. We often refer to such submatrices as monochromatic rectangles.

The following is an easy consequence of Lemma 2.2.13.

Lemma 2.6.2. *Let (U,V) be a MV family in $\mathbb{Z}_m^n$ of size $t \geq 3m$ and let $\omega = exp(2\pi i/m)$ be a primitive root of unity of order m. Then there exists some $1 \leq j \leq m-1$ such that*

$$D_{\omega^j}(U,V) \geq \frac{2}{3m^{3/2}}.$$

Proof. Let μ be the random variable which chooses $u \in U$ and $v \in V$ randomly and outputs $\langle u, v\rangle$ and let $\mathbb{U}_m$ be the uniform distribution over $\mathbb{Z}_m$. Now, $\Delta(\mu, \mathbb{U}_m) \geq (1/2)\left(\mathbf{Pr}[\mathbb{U}_m = 0] - \mathbf{Pr}[\mu = 0]\right) = (1/2)(1/m - 1/t) \geq 1/3m$ as $t \geq 3m$. By Lemma 2.2.13, for some $1 \leq j \leq m-1$,

$$\left|\mathbb{E}_{x\sim\mu}\left[\left(\omega^j\right)^x\right]\right| \geq \frac{2}{3m^{3/2}}.$$

Thus, we have $\left|\mathbb{E}_{u\sim U, v\sim V}\left[\left(\omega^j\right)^{\langle u,v\rangle}\right]\right| \geq \frac{2}{3m^{3/2}}$ as claimed. □

An important ingredient in the proof of Theorem 2 is the following lemma, referred to in the introduction as the 'sub-matrix lemma' which is a generalization of a result of [20].

Lemma 2.6.3 (Sub-Matrix Lemma)**.** *Let $s, m, n \geq 2$ where s divides m, and let ω be a primitive root of unity of order s. Let $A, B \subset \mathbb{Z}_s^n$ be two twin-free lists satisfying $D_\omega(A, B) \geq \frac{2}{3m^{3/2}}$. Let $\operatorname{rank}(P_{A,B}) = r \geq 2$. Then assuming Conjecture 1 (PFR conjecture), there exist lists $A' \subseteq A, B' \subseteq B$ such that $D_\omega(A', B') = 1$, where $|A'| \geq 2^{-c(m)r/\log r}|A|$, $|B'| \geq 2^{-c(m)r/\log r}|B|$ for some constant $c(m)$ which depends only on m.*

Without loss of generality, we can assume $c(m) \geq 1$ above (it will be convenient to assume it in the proof of Theorem 2). In other words, we can replace the $c(m)$ above by $\max\{c(m), 1\}$. We postpone the proof of Lemma 2.6.3 to Section 2.7 and proceed now with the proof of Theorem 2.

We restate Theorem 2 here for convenience and with the explicit function $d(m)$.

Theorem 2.6.4. *Let $n, m \geq 2$ be arbitrary positive integers. Then, assuming Conjecture 1 (PFR conjecture), we have*

$$\mathbf{MV}(m, n) < 2^{d(m)n/\log n},$$

where $d(m) = 1200c(m)m^{6\log m}$ and $c(m)$ is as in Lemma 2.6.3.

Proof. We prove the theorem by induction on the number of (not necessarily distinct) prime factors of m.

Choice of $d(m)$. Let $d, d_1, d_2, d_3 : \mathbb{Z}^+ \to \mathbb{R}$ be functions and d_4 be a constant. We want the following conditions to be satisfied for all $m, n \geq 2$.

1. $d(m), d_1(m), d_2(m), d_3(m)$ are monotonically increasing in m
2. $(2n)^m \leq 2^{d(m)n/\log n}$
3. $(2m)^m \leq 2^{d(m)n/\log n}$
4. $d(m) \geq d(m/2) \cdot 4m \log m$
5. $-d_2(m) + (1/2)\, d(m) > d(m/2) \log m$
6. $2^{(1/2)d(m)n/\log n} \geq 3m2^{d_2(m)n/\log n}$
7. $d_2(m)\, n/\log n \geq 2 \log m + d_3(m)\, n/\log n$
8. $d_3(m) \geq d_1(m) \cdot d_4 \cdot m \log m$
9. $d_4 \geq 300$
10. $d_1(m) \geq 2c(m)$
11. $d_2 \geq d_3 + 1$

It can be verified that the following choice for the functions meets the above conditions.

- $d(m) = 1200 \cdot c(m) \cdot m^{6 \log m}$
- $d_1(m) = 2 \cdot c(m)$

- $d_2(m) = 602 \cdot c(m) \cdot m \log m$
- $d_3(m) = 600 \cdot c(m) \cdot m \log m$
- $d_4 = 300$

We shall explicitly mention which conditions of the above functions are being used in different parts of the proof.

Base Case. The base case is where $m = p$ is prime. Lemma 2.2.9 implies that $\mathbf{MV}(p, n) \leq 1 + \binom{n+p-2}{p-1} < (2\max\{n, p\})^p$. If we show $(2n)^p \leq 2^{d(p)n/\log n}$ and $(2p)^p \leq 2^{d(p)n/\log n}$ we will be done. Indeed, by the choice of $d(m)$ (Condition 2 and 3) both of the above will hold.

Inductive Case. Let $n \geq 2, m \geq 2$ be arbitrary positive integers. Suppose, by induction, that $\mathbf{MV}(s, n) < 2^{d(s)n/\log n}$ for all $s|m, s < m$. We need to show that, assuming Conjecture 1,

$$\mathbf{MV}(m, n) < 2^{d(m)n/\log n}$$

Suppose not. That is, there exists a matching vector family (U, V) of size $t \geq 2^{d(m)n/\log n}$. First, we shall apply Lemma 2.5.4 to (U, V) to obtain a large enough collision free matching vector family (U', V').

A large collision free matching vector family. We show that $|(U', V')| \geq 2^{(1/2)d(m)n/\log n}$. Let $|(U', V')| = t'$. Observe that by Lemma 2.5.4, the inductive hypothesis and the monotonicity of $d(m)$ (Condition 1), $t' \geq 2^{d(m)n/\log n - 2m \cdot d(m/2) \cdot n \log m/\log n}$

where we have used a loose upper bound of m for the number of factors of m. Now,

$$\begin{aligned} & t' \geq 2^{(1/2)d(m)n/\log n} \\ if \quad & d(m)\, n/\log n - 2m \cdot d(m/2) \cdot n \log m/\log n \geq (1/2)\, d(m)\, n/\log n \\ \Leftrightarrow \quad & d(m) \geq d(m/2) \cdot 4m \log m \end{aligned}$$

which is satisfied by the choice of $d(m)$ (Condition 4).

Two key claims. We will need two claims from which the inductive claim follows easily. We shall provide proofs to these claims after the proof of the inductive claim.

Claim 2.6.5. *Let (U, V) be a collision free matching vector family in $\mathbb{Z}_m^n$ with $|(U,V)| \geq 3m$ and $\text{colrank}\left(P_{U,V}^{(s')}\right) > 2$ for all $s'|m, s' \geq 2$. Then, for some $s|m, s \geq 2$, there exists a collision free matching vector family $(U', V') \subseteq (U, V)$ in $\mathbb{Z}_m^n$ satisfying*

- $|(U', V')| \geq 2^{-d_1(m)r_s/\log r_s} |(U, V)|$ *where* $r_s = \text{rank}\left(P_{U,V}^{(s)}\right)$.
- *Either* $\text{colrank}\left(P_{U',V'}^{(s)}\right) \leq (3/4)\,\text{colrank}\left(P_{U,V}^{(s)}\right)$ *or* $\text{colrank}\left(P_{U',V'}^{(s)}\right) \leq 2$.

Claim 2.6.6. *Let (U, V) be a collision free matching vector family in $\mathbb{Z}_m^n$ such that $|(U,V)| \geq 3m \cdot 2^{d_2(m)n/\log n}$. Then, there exists a collision free matching vector family $(U', V') \subseteq (U, V)$ in $\mathbb{Z}_m^n$ satisfying*

- $|(U', V')| \geq 2^{-d_2(m)n/\log n} |(U, V)|$.
- $P_{U,V}^{(s)}$ *is the all zero matrix for some* $s|m, s \geq 2$.

Let us proceed with the proof of the inductive claim assuming these two claims. We have a collision free matching vector family (U', V') with $|(U', V')| \geq 2^{(1/2)d(m)n/\log n} \geq 3m \cdot 2^{d_2(m)n/\log n}$. (Condition 6 satisfied by the choice of $d(m), d_2(m)$) Applying Claim 2.6.6, there exists a collision free matching vector family $(U'', V'') \subseteq (U', V') \subseteq (U, V)$ in $\mathbb{Z}_m^n$ satisfying

- $|(U'', V'')| \geq 2^{-d_2(m)n/\log n} 2^{(1/2)d(m)n/\log n}$.
- $P^{(s)}_{U'',V''}$ is the all zero matrix for some $s|m, s \geq 2$.

By the choice of $d(m)$, it can be verified that $-d_2(m) + (1/2)\, d(m) > d(m/2) \log m$ (Condition 5). Thus, $|(U'', V'')| > 2^{d(m/2)n \log m/\log n}$.

We now show that this is enough to get a contradiction. If $s = m$, we have $|(U'', V'')| \leq 1$ as (U'', V'') is a matching vector family in $\mathbb{Z}_m^n$. If $s < m$, by Lemma 2.5.2 and the inductive hypothesis, we have $|(U'', V'')| \leq 2^{d(m/s)n \log m/\log(n \log m)} \leq 2^{d(m/2)n \log m/\log n}$ by monotonicity of $d(m)$ (Condition 1). Thus, irrespective of s, $|(U'', V'')| \leq 2^{d(m/2)n \log m/\log n}$ which is a contradiction. This completes the proof. □

Proof of Claim 2.6.5: Let $|(U, V)| = t \geq 3m$. Let ω be a root of unity of order m. By Lemma 2.6.2, for some $1 \leq j \leq m-1$, $D_{\omega^j}(U, V) \geq \frac{2}{3m^{3/2}}$. Note that $s = m/gcd(m, j)$ is the order of $\omega' = \omega^j$. Observe that $s|m, s \geq 2$ as $1 \leq j \leq m-1$. Recall from the statement of the claim that $r_s = rank\left(P^{(s)}_{U,V}\right)$. Thus, by the collision

free property of (U, V),

$$D_{\omega'}\left(U^{(s)}, V^{(s)}\right) = \left|\mathbb{E}_{u\sim U^{(s)}, v\sim V^{(s)}}\left[(\omega')^{\langle u,v\rangle}\right]\right| = \left|\mathbb{E}_{u\sim U, v\sim V}\left[(\omega')^{\langle u,v\rangle}\right]\right| = D_{\omega'}(U, V) \geq \frac{2}{3m^{3/2}}.$$

Applying Lemma 2.6.3 on $U^{(s)}, V^{(s)}$ with ω' a primitive root of unity of order s, we can get an $(R \times S)$ submatrix of $P_{U,V}$ with $|R| = |S| \geq 2^{-c(m)r_s/\log r_s}t$. (we can make $|R| = |S|$ as throwing away rows and columns from a monochromatic rectangle still keeps it monochromatic) Let $T = R \cap S$. We divide our analysis to two cases: either $|T| > |R|/2$ or $|T| \leq |R|/2$. In both cases, we shall exhibit a matching vector family as required in the statement of the claim.

Case 1: $|T| > |R|/2$. For $U = (u_1, u_2, \cdots u_t)$, $V = (v_1, v_2, \cdots v_t)$, let $U' = (u_j | j \in T)$ and $V' = (v_j | j \in T)$, and $P' = P_{U',V'}$. Now, as $P'^{(s)}$ is monochromatic, and $\langle u_j, v_j\rangle = 0 \pmod s$ for $j \in T$, we have $\langle u', v'\rangle = 0 \,(mod\ s)$ for all $u' \in U', v' \in V'$. Observe that

- $|(U', V')| \geq 2^{-1-c(m)r_s/\log r_s}t \geq 2^{-2c(m)r_s/\log r_s}t \geq 2^{-d_1(m)r_s/\log r_s}t$ (by the choice of $d_1(m)$, Condition 10)
- $\text{colrank}\left(P^{(s)}_{U',V'}\right) = 0 \leq 2$

This finishes Case 1.

Case 2: $|T| \leq |R|/2$. Let $R' = R \setminus T$ and $S' = S \setminus T$. Note that $R' \cap S' = \emptyset$ and $|R'| = |S'|$. Consider the $R' \cup S' \times R' \cup S'$ submatrix of $P_{U,V}$. Call it P'. Note that

$$P'^{(s)} = \begin{pmatrix} P'_1 & C \\ \star & P'_2 \end{pmatrix}$$

where P'_1 and P'_2 are the $R' \times R'$ and the $S' \times S'$ submatrices of $P^{(s)}_{U,V}$ respectively and C is monochromatic. We add a matrix of column rank at most 1 to $P'^{(s)}$ to yield $P''^{(s)}$ which is the same as $P'^{(s)}$ except that C is replaced by the all zero block matrix. Thus,

$$P''^{(s)} = \begin{pmatrix} P'_1 & 0 \\ \star & P'_2 \end{pmatrix}$$

Note that by Claim 2.4.8, $\text{colrank}\left(P''^{(s)}\right) \leq \text{colrank}\left(P'^{(s)}\right) + 1$. Now, using Claim 2.4.9, $\text{colrank}\left(P'_1\right) + \text{colrank}\left(P'_2\right) \leq \text{colrank}\left(P'^{(s)}\right) + 1 \leq \text{colrank}\left(P^{(s)}_{U,V}\right) + 1 \leq (3/2)\,\text{colrank}\left(P^{(s)}_{U,V}\right)$ as $\text{colrank}\left(P^{(s)}_{U,V}\right) > 2$. Therefore, one of P'_1, P'_2, say P'_1 satisfies $\text{colrank}\left(P'_1\right) \leq (3/4)\,\text{colrank}\left(P^{(s)}_{U,V}\right)$. Construct the matching vector family (U', V') as follows. Let $U' = (u_j | j \in R')$ and $V' = (v_j | j \in R')$. Again, observe that

- $|(U', V')| \geq 2^{-1-c(m)r_s/\log r_s} t \geq 2^{-2c(m)r_s/\log r_s} t \geq 2^{-d_1(m)r_s/\log r_s} t$ (by the choice of $d_1(m)$, Condition 10).
- $\text{colrank}\left(P^{(s)}_{U',V'}\right) \leq (3/4)\text{colrank}\left(P^{(s)}_{U,V}\right)$.

This completes the proof of Case 2.

□

Proof of Claim 2.6.6: We will use Claim 2.6.5 iteratively. For this, we first set up some notations.

The setup. Define a sequence of collision free matching vector families for $i = 0, \ldots, z$.

- $(U, V) = (U_0, V_0), (U_1, V_1) \cdots$

- Let $t_i = |(U_i, V_i)|$.
- Each step i has label $s_i|m$ (this label will be given by Claim 2.6.5).
- Let $cr_i : \mathbb{Z}^+ \to \mathbb{R}$ be defined by

$$cr_i(s) = \text{colrank}\left(P^{(s)}_{U_i,V_i}\right).$$

- Let $r_i : \mathbb{Z}^+ \to \mathbb{Z}$ be defined by

$$r_i(s) = \text{rank}\left(P^{(s)}_{U_i,V_i}\right).$$

Invariants. We will show how to go from step i to step $i+1$. We stop after stage z when $cr_z(s) \leq 2$ for some $s|m, s \geq 2$. We shall maintain the following invariants for $0 \leq i \leq z-1$.

- $(U_{i+1}, V_{i+1}) \subseteq (U_i, V_i)$ and hence is a collision free matching vector family in $\mathbb{Z}_m^n$.
- $t_{i+1} \geq 2^{-d_1(m)r_i(s_i)/\log r_i(s_i)} t_i$.
- $cr_{i+1}(s_i) \leq (3/4)\, cr_i(s_i)$ or $cr_{i+1}(s_i) \leq 2$.
- $cr_{i+1}(s') \leq cr_i(s')$ for all $s'|m$.

Step $i \to$ Step $i+1$. We state a claim that we will prove below.

Claim 2.6.7. $\sum_{i=0}^{z-1} d_1(m)\, r_i(s_i) / \log r_i(s_i) \leq d_3(m)\, n/logn$.

In order to apply Claim 2.6.5, we need to satisfy $t_i \geq 3m$. Observe that by Claim 2.6.7,

$$\begin{aligned} t_i \geq t_z &\geq t_0 \prod_{j=0}^{z-1} 2^{-d_1(m) r_j(s_j)/\log r_j(s_j)} \\ &\geq 2^{-d_3(m)n/\log n} t_0 \\ &\geq 3m \cdot 2^{-d_3(m)n/\log n + d_2(m)n/\log n} \geq 3m, \end{aligned}$$

(by the choice of $d_2(m), d_3(m)$ in Condition 11). Apply Claim 2.6.5 to (U_i, V_i) to get label s_i for step i and $(U_{i+1}, V_{i+1}) \subseteq (U_i, V_i)$. The first three invariants are maintained by the statement of Claim 2.6.5. The last invariant follows from Fact 2.4.4. Note that by the inequality we just established, $t_z \geq 2^{-d_3(m)n/\log n} t_0$. Also, by the stopping condition, $cr_z(s') \leq 2$ for some $s'|m, s' \geq 2$. Thus, applying Claim 2.4.10, we get another matching vector family $(U', V') \subseteq (U_z, V_z) \subseteq (U, V)$ such that

- $|(U', V')| \geq t_z/m^2 \geq 2^{-2\log m - d_3(m)n/\log n} |(U, V)| \geq 2^{-d_2(m)n/\log n} |(U, V)|$ (Condition 7 satisfied by the choice of $d_2(m)$ and $d_3(m)$).
- $P^{(s')}_{U', V'}$ is the all zero matrix.

This finishes the Proof of Claim 2.6.6.

Proof of Claim 2.6.7: Let t_s be the number of steps with label s. Note that as the column rank modulo s goes down by a factor of at least 3/4 each time we are in a step labeled s, it is easy to see that $t_s \leq \log_{4/3} cr_0(s) \leq \log_{4/3} n$. We shall rely on the monotonic increasing nature of $x/\log x$ when $x \geq e$. As $cr_i(s) > 2$, by Claim 2.4.6, $r_i(s) \geq cr_i(s) > 2$ which means $r_i(s) \geq 3 > e$ as the rank is always an

integer. We thus have

$$
\begin{aligned}
& \sum_{i=0}^{z-1} d_1(m) \frac{r_i(s_i)}{\log r_i(s_i)} \\
\leq\ & d_1(m) \log m \sum_{i=0}^{z-1} \frac{cr_i(s_i)}{\log cr_i(s_i)} \quad \text{(by Claim 2.4.6) and monotonicity of } x/\log x \text{ as discussed above} \\
\leq\ & d_1(m) \log m \sum_{s|m, s\geq 2} \sum_{j=1}^{\lfloor \log_{4/3} n(s) \rfloor} \left(\frac{cr_0(s)}{(4/3)^{j-1} \log\left(cr_0(s)/(4/3)^{j-1}\right)} \right) \\
\leq\ & d_1(m) \log m \sum_{s|m, s\geq 2} d_4 cr_0(s)/\log cr_0(s) \quad \text{(Condition 9 satisfied by } d_4) \\
\leq\ & d_1(m) \log m \sum_{s|m, s\geq 2} d_4 n/\log n \quad \text{(as } cr_0(s) \leq r_0(s) \leq r_0(m) \leq n, \text{ by Claim 2.4.6 and Fact 2.4.5)} \\
\leq\ & d_4 d_1(m)\, m (\log m)\, n/\log n \\
\leq\ & d_3(m)\, n/\log n \quad \text{(by the choice of } d_3(m), \text{ Condition 8)}
\end{aligned}
$$

This completes the proof. □

2.7 Monochromatic rectangles from low rank matrices

In this section we prove Lemma 2.6.3 (the Sub-Matrix Lemma). We begin with some preliminary definitions. The following is a standard result in algebra and can be find in any introductory text.

Theorem 2.7.1 (Fundamental Theorem of finitely generated abelian groups)**.** *Every finitely generated abelian group G is isomorphic to a direct product of cyclic groups of prime power order and an infinite cyclic group. More precisely,*

$$G \cong \mathbb{Z}^n \times \mathbb{Z}_{q_1} \times \mathbb{Z}_{q_2} \cdots \times \mathbb{Z}_{q_r}$$

where q_i's are prime powers with $q_1 \leq q_2 \cdots \leq q_r$. The decomposition is unique after applying this ordering on q_i's. If the group G is finite, then $n = 0$.

We will use the following two definitions regarding sumsets.

Definition 2.7.2 (Difference Set)**.** *For $A \subseteq \mathbb{Z}_m^n$ define its difference set as $A - A = \{a - a' | a, a' \in A\}$.*

Definition 2.7.3 ($rep_S(x)$)**.** *For any $S \subseteq \mathbb{Z}_m^n$ and $x \in \mathbb{Z}_m^n$, $rep_S(x)$ is the number of different representations of x as an expression of the form $s - s'$ where $s, s' \in S$.*

Next, we define the ϵ-spectrum of B with respect to a primitive root of unity of order m.

Definition 2.7.4 (Spectrum)**.** *For $B \subseteq \mathbb{Z}_m^n$, and $\epsilon \in [0, 1]$, the ϵ-spectrum of B with respect to ω, a primitive root of unity of order m, is the set*

$$\mathrm{Spec}_\epsilon(B) = \left\{ x \in \mathbb{Z}_m^n : \left| \mathbb{E}_{b \sim B} \left[\omega^{\langle x, b \rangle} \right] \right| \geq \epsilon \right\}.$$

When ω is implicit in the context, we will drop the phrase "with respect to ω".

We start by proving the following lemma which is a generalization of a lemma from [20].

Lemma 2.7.5. *Let $A, B \subseteq \mathbb{Z}_m^n$ be sets. Let ω be a primitive root of unity of order m. If $A \subseteq \mathrm{Spec}_\epsilon(B)$, then there exist sets $A' \subseteq A, B' \subseteq B$, such that $|A'| \geq |A|/m$ and $|B'| \geq \epsilon^2 \frac{|A|}{|span(A)|} |B|$ such that $D_\omega(A', B') = 1$.*

Proof. We start by setting up some notations. Let $W = \text{span}\,(A)$ be the subgroup of $\mathbb{Z}_m^n$ spanned by A. By Theorem 2.7.1, there exists an isomorphism $\tau : \prod_{i=1}^r \mathbb{Z}_{q_i} \to W$. Let $\mathbb{C} = \prod_{i=1}^r \mathbb{Z}_{q_i}$ and note that we can think of elements of $\mathbb{C}$ as vectors with integer coordinates where the i'th coordinate is in $\mathbb{Z}_{q_i}$. Let $e_1, e_2, \cdots e_r \in \mathbb{C}$ where e_i is the vector that has 1 in the i'th coordinate and 0 everywhere else. Given $x \in \mathbb{C}$, $\exists \alpha_1, \cdots \alpha_r$, with $\alpha_i \in \mathbb{Z}_{q_i}$ such that

$$x = \sum_{i=1}^{r} \alpha_i e_i.$$

Then $\tau\,(x) = \sum_{i=1}^r \alpha_i \tau\,(e_i)$. Let $v_i = \tau\,(e_i)$ for $1 \leq i \leq r$. We can think of the v_i's as a basis of W. Therefore, for $\alpha = (\alpha_1, \alpha_2, \cdots \alpha_r) \in \mathbb{C}$ we have $\tau\,(\alpha) = \sum_{i=1}^r \alpha_i v_i$. Let

$$\Theta = \{(\beta_1, \cdots \beta_r) \in \mathbb{Z}_m^r | \exists u \in \mathbb{Z}_m^n \text{ such that } \forall i, \beta_i = \langle v_i, u \rangle\}.$$

Claim 2.7.6. *For $1 \leq i \leq r$, $q_i v_i = 0^n \pmod m$.*

Proof. Let $x = 0^r \in \mathbb{C}$. Now $\tau\,(x) = 0^n \pmod m$. Note that x can also be written as $x = q_i e_i$. Applying τ on both sides, we get $\tau\,(x) = q_i v_i$. Thus, $q_i v_i = 0^n \pmod m$. □

Claim 2.7.7. *For $\beta \in \Theta$, $1 \leq i \leq r$, $q_i \beta_i = 0 \pmod m$.*

Proof. As $\beta \in \Theta$, there is a $u \in \mathbb{Z}_m^n$ such that $\forall i$, $\beta_i = \langle v_i, u \rangle$. Then, $q_i \beta_i = q_i \langle v_i, u \rangle = 0 \pmod m$ by Claim 2.7.6. □

For $\alpha \in \mathbb{C}, \beta \in \Theta$ we define their inner product $\langle \alpha, \beta \rangle \in \mathbb{Z}_m$ by considering $\alpha_i \in \{0, \ldots, q_i - 1\}, \beta_i \in \{0, \ldots, m-1\}$, taking the inner product over the inte-

gers and then reducing the result modulo m. This is indeed an inner product by Claim 2.7.7.

Claim 2.7.8. *Given* $\beta \in \Theta \setminus \{0\}$,

$$\sum_{a \in W} \omega^{\langle \tau^{-1}(a), \beta \rangle} = \sum_{\alpha \in \mathbb{C}} \omega^{\langle \alpha, \beta \rangle} = 0.$$

Proof. Let $\beta_i \neq 0$. Then $\sum_{\alpha \in C} \omega^{\langle \alpha, \beta \rangle} = 0$ whenever $\sum_{j=0}^{q_i - 1} \omega^{j\beta_i} = 0$. Now, $\sum_{j=0}^{q_i-1} \omega^{j\beta_i} = \frac{\omega^{q_i \beta_i} - 1}{\omega^{\beta_i} - 1}$. This is well defined because ω is of order m and $\beta_i \neq 0$. The claim now follows from Claim 2.7.7 which makes the expression zero. □

With the above setup in place, we can now proceed with the proof of Lemma 2.7.5. For $\beta \in \Theta$, define

$$S_\beta = \{x \in \mathbb{Z}_m^n | \langle v_i, x \rangle = \beta_i, 1 \leq i \leq r\}.$$

Denoting $\mu(\beta) = \mathbf{Pr}_{b \in B}[b \in S_\beta]$, we observe that $\cup_{\beta \in \Theta} (B \cap S_\beta) = B$. Hence, $\sum_{\beta \in \Theta} \mu(\beta) = 1$. For $a \in W$, define $h(a) = \mathbb{E}_{b \in B}\left[\omega^{\langle a, b \rangle}\right]$. If $a = \sum_{i=1}^r \alpha_i v_i$ then

$$\begin{aligned} h(a) &= \mathbb{E}_{b \in B}\left[\omega^{\langle a, b \rangle}\right] \\ &= \mathbb{E}_{b \in B}\left[\omega^{\langle \sum_{i=1}^r \alpha_i v_i, b \rangle}\right] \\ &= \sum_{\beta \in \Theta} \mu(\beta) \omega^{\langle \alpha, \beta \rangle} \\ &= \sum_{\beta \in \Theta} \mu(\beta) \omega^{\langle \tau^{-1}(a), \beta \rangle}. \end{aligned}$$

We will prove upper and lower bounds for the sum $\sum_{a\in A}|h(a)|^2$. On the one hand,

$$\begin{aligned}\sum_{a\in A}|h(a)|^2 &\geq \frac{1}{|A|}\left(\sum_{a\in A}|h(a)|\right)^2 \text{ (Cauchy Scwartz inequality)}\\ &\geq \frac{1}{|A|}\left(\sum_{a\in A}\epsilon\right)^2 \ (A\subseteq \mathrm{Spec}_\epsilon(B) \text{ implies } |h(a)|\geq\epsilon\ )\\ &\geq |A|\epsilon^2.\end{aligned}$$

On the other hand,

$$\begin{aligned}\sum_{a\in A}|h(a)|^2 &\leq \sum_{a\in W}|h(a)|^2\\ &= \sum_{a\in W}\sum_{\beta\in\Theta,\beta'\in\Theta'}\mu(\beta)\mu(\beta')\omega^{\langle\tau^{-1}(a),\beta-\beta'\rangle}\\ &= \sum_{\beta,\beta'\in\Theta,}\mu(\beta)\mu(\beta')\sum_{a\in W}\omega^{\langle\tau^{-1}(a),\beta-\beta'\rangle}\\ &= \sum_{\beta\in\Theta}\mu(\beta)^2|W| \qquad \text{(Claim 2.7.8)}\\ &\leq |W|\max_{\beta\in\Theta}\{\mu(\beta)\}.\end{aligned}$$

Now, combining the upper and lower bounds, $\max_{\beta\in\Theta}\{\mu(\beta)\}\geq\epsilon^2\frac{|A|}{|W|}$. Thus, there exists a $\beta\in\Theta$ such that $\mu(\beta)\geq\epsilon^2\frac{|A|}{|W|}$. This means that the subset $B'=B\cap S_\beta$ is of size at least $\epsilon^2\frac{|A|}{|W|}|B|$. Now, any $a\in A$ can be written as $a=\sum_{i=1}^r\alpha_i v_i$, and for $b\in B'$, the inner product $\langle a,b\rangle=\langle\alpha,\beta\rangle$ is independent of b. Now, for $i\in[m]$, let $A_i\subseteq A$ be such that for $a\in A_i$, for all $b\in B'$, $\langle a,b\rangle=\langle\alpha,\beta\rangle=i$. Now there exists some A_i, call it A', of size at least $|A|/m$ such that $\langle a',b'\rangle=i$ for all $a'\in A',b'\in B'$, that is, $D_\omega(A',B')=1$ and this proves the lemma. □

We shall also need the following simple lemma. In the following, for a set B, let $arc(a)$ be the phase of the complex number $\mathbb{E}_{b\in B}\omega^{\langle a,b\rangle}$.

Claim 2.7.9. *Let $A \subseteq Spec_\epsilon(B)$. Let $s \in \{0,1,2,3\}$ be the integer that maximizes the size of the multiset*

$$\{arc(a) \cap [s\pi/2, (s+1)\pi/2) : a \in A\}$$

Now, set

$$A' = \{a \in A : \{arc(a) \in [s\pi/2, (s+1)\pi/2)\}$$

Then, $D_\omega(A', B) \geq \epsilon/4$.

Proof. We prove it for $s = 0$. The proof for any other s is exactly similar. Then $\mathbb{E}_{a\in A', b\in B}[\omega_{\langle a,b\rangle}] = x + iy$ where $x > \epsilon\ \mathbb{E}_{a\in A'}[cos\ arc(a)]$ and $y \geq \epsilon\ \mathbb{E}_{a\in A'}[sin\ arc(a)]$. Now we show that $|x + iy| \geq \epsilon/4$ which will prove the lemma. To see this, if $\mathbb{E}a \in A'[cos\ arc(a)] \geq 1/4$ we are done. Now suppose not. Then by Markov inequality, $\mathbf{Pr}_{a\in A'}[cos\ arc(a) > 1/2] < 1/2$. Thus, $\mathbf{Pr}_{a\in A'}[sin\ arc(a) \geq \sqrt{3}/2] > 1/2$. Therefore, $\mathbb{E}a \in A'[sin\ arc(a)] \geq \sqrt{3}/4 > 1/4$. □

We continue along the lines of [20] and prove the following lemma.

Lemma 2.7.10. *Suppose the twin free lists $U, V \subseteq \mathbb{Z}_m^n$ satisfy $D_\omega(U, V) \geq \epsilon$ where ω is a primitive root of unity of order m. Also, let $\operatorname{rank}(P_{U,V}) = r$. Then assuming Conjecture 1, for every $K > 1$, letting $\ell = r/\log_m K$, there exist lists $U' \subseteq U, V' \subseteq V$ such that $D_\omega(U', V') = 1$, and $|U'| \geq poly_m\left(\frac{(\epsilon/2)^{2^\ell}}{rK}\right)(2mr)^{-\ell}|U|$, $|V'| \geq poly_m\left(\frac{(\epsilon/2)^{2^\ell}}{rK}\right)m^{-\ell}|V|$.*

Proof. Let $U = (u_1, \cdots u_t)$ and $V = (v_1, \cdots v_t)$. Since $P_{U,V}$ has rank r there exists a $t \times r$ matrix U_M and $r \times t$ matrix V_M so that $U_M V_M = P_{U,V}$. Thus if we let A

denote the rows of U_M and B denote the columns of V_M, then $A, B \subseteq \mathbb{Z}_m^r$. The proof does not care about the order of elements and hence we now consider A, B which are sets. Note that $|A| = |B| = t$ and if $A = (a_1, \cdots a_t)$ and $B = (b_1, \cdots b_t)$ then $\langle a_i, b_j \rangle = \langle u_i, v_j \rangle$ for $1 \leq i, j \leq t$. Thus, $D_\omega(U, V) \geq \epsilon$ implies $D_\omega(A, B) \geq \epsilon$. Following [20] consider a sequence of constants $\epsilon_1 = \epsilon/8$, $\epsilon_2 = \epsilon_1^2/8$, $\epsilon_3 = \epsilon_2^2/8$, $\cdots$ and a sequence of sets $A_1' = A \cap \mathrm{Spec}_{\epsilon_1}(B)$ and $A_i' \subseteq (A_{i-1} - A_{i-1}) \cap \mathrm{Spec}_{\epsilon_i}(B)$ and a sequence of sets $A_i \subseteq A_i'$. The way the subsets are chosen will be made precise shortly. Now by the pigeonhole principle, there exists a minimal index $\ell \leq r/\log_m K$ such that $|A_{\ell+1}| \leq K|A_\ell|$. and let

$$A_i' = \{a - a' | a, a' \in A_{i-1}, a - a' \in \mathrm{Spec}_{4\epsilon_i}(B) \ \textit{and}\ m^{j_i} \leq rep_{A_{i-1}}(a - a') \leq m^{j_i+1}\}.$$

By Claim 2.7.9, $D_\omega(A_i, B) \geq \epsilon_i$.

Claim 2.7.11. *For $i = 1$ we have $|A_1| \geq (\epsilon_1)|A|$. For $i > 1$ we have $\mathbf{Pr}_{a,a' \in A_{i-1}}[a - a' \in A_i'] \geq 4\epsilon_i/r$ and additionally $|A_i| \geq \frac{\epsilon_i}{m^{j_i+1} r}|A_{i-1}|^2$.*

Proof. If $i = 1$, by Markov inequality, $|A_1'| \geq \epsilon_2|A|$. Therefore, $|A_1| \geq |A_1'|/4 \geq \epsilon_1|A|$. For larger i, we show that

$$\mathbf{Pr}_{a,a' \in A_{i-1}}[a - a' \in \mathrm{Spec}_{\epsilon_i}(B)] \geq \epsilon_i.$$

This follows from the fact that

$$\epsilon_{i-1}^2 \leq \left|\mathbb{E}_{b \in B, a \in A_{i-1}}\left[\omega^{\langle a,b \rangle}\right]\right|^2 \leq \mathbb{E}_{b \in B}\left|\mathbb{E}_{a \in A_{i-1}}\left[\omega^{\langle a,b \rangle}\right]\right|^2 = \mathbb{E}_{a,a' \in A_{i-1}}\mathbb{E}_{b \in B}\left[\omega^{\langle a-a',b \rangle}\right].$$

Now applying Markov inequality we get that $\mathbf{Pr}_{a,a' \in A_{i-1}}[a - a' \in \mathrm{Spec}_{\epsilon_i}(B)] \geq 4\epsilon_i = \epsilon_{i-1}^2/2$. Now selecting j_i as in the construction gives that $\mathbf{Pr}_{a,a' \in A_{i-1}}[a - a' \in A_i'] \geq 4\epsilon_i/r$.

To prove the second part of the lemma, observe that by the above, we have shown that

$$|\{(a,a') \in A_{i-1} \times A_{i-1} | a - a' \in A_i\}| \geq \frac{4\epsilon_i}{r}|A_{i-1}|^2.$$

Also, by construction This completes the proof. □

Below we will use the following additive-combinatorics lemma.

Theorem 2.7.12 ([10, 79])**.** *There exists an absolute constant $c > 0$ such that the following holds. Let A be any arbitrary subset of an abelian group G. Let $S \subseteq G$ be such that $|S| \leq C|A|$. If $\mathbf{Pr}_{a,a' \in A}[a - a' \in S] \geq 1/C$, then there exists a subset $A' \subseteq A$ such that $|A'| \geq \frac{|A|}{C^c}$ and $|A' - A'| \leq C^c|A|$.*

Now we come to the main claim.

Claim 2.7.13. *For $i = \ell, \ell - 1, \cdots 1$ there exist subsets $A'_i \subseteq A_i$, $B'_i \subseteq B$ such that $D_\omega(A'_i, B'_i) = 1$ and*

$$|A'_i| \geq \alpha_i |A_i|$$

and

$$|B'_i| \geq \beta_i |B|$$

where $\alpha_i = poly_m\left(\frac{\epsilon_{\ell+1}}{rK}\right)(2mr)^{-(\ell-i)}\left(\prod_{j=i}^{\ell}\epsilon_{j+1}\right)$, $\beta_i = poly_m\left(\frac{\epsilon_{\ell+1}}{rK}\right)m^{-(\ell-i)}$

Base Case. The base case of $i = \ell$ is proved by an application of the Balog-Szemeredi-Gowers theorem followed by Conjecture 1 followed by Lemma 2.7.5. To see this, we know that $|A'_{\ell+1}| \leq 4|A_{\ell+1}| \leq 4K|A_\ell|$ and $\mathbf{Pr}_{a,a' \in A_\ell}[a - a' \in A'_{\ell+1}] \geq 4\epsilon_{\ell+1}/r$. Hence by Theorem 2.7.12 (with $C = \frac{rK}{\epsilon_{\ell+1}}$), there exists a set $A''_\ell \subseteq A_\ell$ such

that $|A''_\ell| \geq poly\left(\frac{\epsilon_{\ell+1}}{rK}\right)|A_\ell|$ and $|A''_\ell - A''_\ell| \leq poly\left(\frac{rK}{\epsilon_{\ell+1}}\right)|A''_\ell|$. Now by Conjecture 1 applied to A''_ℓ, there exists a set $A'''_\ell \subseteq A''_\ell$ such that $|A'''_\ell| \geq poly_m\left(\frac{\epsilon_{\ell+1}}{rK}\right)|A''_\ell|$ and $|\text{span}\,(A'''_\ell)| \leq m|A''_\ell| = poly_m\left(\frac{rK}{\epsilon_{\ell+1}}\right)|A'''_\ell|$. (Note the extra factor of m in front of $|A''_\ell|$ as we get a coset of size $|A''_\ell|$ and its span incurs an additional factor of m) Also, as $A'''_\ell \in \text{Spec}_{\epsilon_\ell}(B)$, applying Lemma 2.7.5 to A'''_ℓ and B, we get $A'_\ell \subseteq A'''_\ell$ and $B'_\ell \subseteq B$ such that $D_\omega(A'_\ell, B'_\ell) = 1$, $|A'_\ell| \geq poly_m\left(\frac{\epsilon_{\ell+1}}{rK}\right)|A_\ell|$ and $|B'_\ell| \geq poly_m\left(\frac{\epsilon_{\ell+1}}{rK}\right)|B|$. This completes the base case. Let us come to the inductive case. □

Inductive Case. Suppose the statement is true for i and let us argue for $i-1$. Let $G = (A_{i-1}, E)$ be the graph whose vertices are the elements in A_{i-1} and (a, a') is an edge if $a - a' \in A'_i$. Now,

$$\begin{aligned}
|E| &\geq m^{j_i}|A'_i| \\
&\geq m^{j_i}\alpha_i|A_i| \quad \text{(inductive hypothesis)} \\
&\geq m^{j_i}\alpha_i\frac{\epsilon_i}{m^{j_i+1}r}|A_{i-1}|^2 \quad \text{(Claim 2.7.11)} \\
&= 2\alpha_{i-1}|A_{i-1}|^2
\end{aligned}$$

Now the graph has at least $2\alpha_{i-1}|A_{i-1}|^2$ edges and $|A_{i-1}|$ vertices and therefore has a vertex $\tilde{a}$ with out degree at least $2\alpha_{i-1}|A_{i-1}|$. Let us call the neighbors vertices A''_{i-1}. Partition B'_i into $B'_{i,j}$ for $0 \leq j \leq m-1$ such that all elements of $B'_{i,j}$ have inner product j with $\tilde{a}$. Let $B'_{i-1} = B_{i,j_1}$ be the largest of them. Note that $|B'_{i-1}| \geq |B'_i|/m$. By assumption $D_\omega(A'_i, B'_i) = 1$. Hence, $D_\omega\left(A'_i, B'_{i-1}\right) = 1$. Therefore, for some j_2, $\langle a, b\rangle = j_2$ for all $a \in A'_i$ and $b \in B'_{i-1}$. Now, we have $(\tilde{a}, a) \in E$ whenever $a \in A''_{i-1}$, and hence $\langle \tilde{a} - a, b\rangle = j_2$ for $b \in B'_{i-1}$. Thus, for $a \in A''_{i-1}$ and $b \in B'_{i-1}$, we have $\langle a, b\rangle = j_1 - j_2$. Thus, $|A'_{i-1}| \geq 2\alpha_{i-1}|A_{i-1}| \geq \alpha_{i-1}|A_{i-1}|$ and

$B'_{i-1} \geq |B'_i|/m \geq \frac{\beta_i}{m}|B| = \beta_{i-1}|B|$. This completes the inductive case. $\square$

Put $i = 1$ in the above claim. Also observe that as $\epsilon_{j+1} = \epsilon^{2^j}/8^{2^{j+1}-1} \geq (\epsilon/64)^{2^j}$. Thus, $\epsilon_{\ell+1} \geq (\epsilon/64)^{2^\ell}$ and $\prod_{j=1}^{\ell} \epsilon_{j+1} \geq (\epsilon/64)^{2^{\ell+1}}$ there exist $A' \subseteq A_1 \subseteq A$, $B' \subseteq B$, such that $|A'| \geq poly\left(\frac{(\epsilon/64)^{2^\ell}}{rK}\right)(2mr)^{-\ell}|A|$ and $|B'| \geq poly\left(\frac{(\epsilon/64)^{2^\ell}}{rK}\right)m^{-\ell}|B|$. Observing that the lower bounds grow weaker with increasing ℓ,and that $\ell \leq \ell' = r/\log_m K$ we get $|A'| \geq poly\left(\frac{(\epsilon/64)^{2^{\ell'}}}{rK}\right)(2mr)^{-\ell'}|A|$ and $|B'| \geq poly\left(\frac{(\epsilon/64)^{2^{\ell'}}}{rK}\right)m^{-\ell'}|B|$ where $\ell' = r/\log_m K$. Therefore, if we take the list $U' \subseteq U$ (corresponding to $A' \subseteq A$) and $V' \subseteq V$ (corresponding to $B' \subseteq B$) then as $\langle a_i, b_j\rangle = \langle u_i, v_j\rangle$ the statement of the lemma follows. This completes the proof of Lemma 2.7.10 $\square$

We can now prove the Sub-Matrix Lemma, Lemma 2.6.3.

Proof of Lemma 2.6.3: Set $K = s^{4r/\log r}, \ell = \frac{\log r}{4}, \epsilon = 1/2m^{3/2}$ while applying Lemma 2.7.10 over $\mathbb{Z}_s$. We get $|A'| \geq \delta_s|A|$, $|B'| \geq \delta_s|B|$ where

$$\begin{aligned}\delta_s &= poly_s\left(\frac{1}{m^{r^{1/4}}}\right)2^{-c_1(s)r/\log r} \quad \text{(for some constant } c_1(s) \text{ depending only on } s\text{)} \\ &\geq poly_m\left(\frac{1}{m^{r^{1/4}}}\right)2^{-c_1(s)r/\log r}\end{aligned}$$

Now let $c_2(m) = \max_{s|m, s\geq 2}\{c_1(s)\}$. Thus, $\delta_s \geq poly_m\left(\frac{1}{m^{r^{1/4}}}\right)2^{-c_2(m)r/\log r} \geq 2^{-c(m)r/\log r}$ for some constant c that depends only on m. $\square$

Chapter 3

A Barrier in Polynomial Lower Bounds

3.1 Introduction

Polynomials play a fundamental role in computer science with important applications in algorithm design, coding theory, pseudo-randomness, cryptography and complexity theory. They are also instrumental in proving lower bounds, as many lower bounds techniques first reduce the computational model to a computation or an approximation by a low degree polynomial, and then continue to show that certain hard functions cannot be computed or approximated by low degree polynomials. Motivated by these applications, the problem of constructing explicit functions which cannot be computed or approximated (in certain ways) by low degree polynomials has been widely explored in computational complexity. However, most techniques to date apply only to relative low degree polynomials. In this paper, we focus on understanding this phenomenon, when the polynomials are defined over fixed size finite fields. In this regime, many lower bound techniques become trivial when the degree grows beyond logarithmic in the number of variables. We propose a new barrier explaining the lack of ability to prove strong lower bounds

for polynomials of super-logarithmic degree. The barrier is based on *nonclassical polynomials*, an extension of standard (classical) polynomials which arose in higher order Fourier analysis. We show that several existing lower bound techniques extend to nonclassical polynomials, for which the logarithmic degree bound is tight. Hence, to prove stronger lower bounds, one should either focus on techniques which distinguish classical from nonclassical polynomials, or consider functions which are hard also for nonclassical polynomials.

Nonclassical polynomials. Nonclassical polynomials were introduced by Tao and Ziegler [159] in their works on the inverse theorem for the Gowers uniformity norms. To introduce these, it will be beneficial to first consider classical polynomials. Fix a prime finite field $\mathbb{F}_p$, where we consider p to be a constant. A function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ is a degree d polynomial if it can be written as a linear combination of monomials of degree at most d. An equivalent definition is that f is annihilated by taking any $d+1$ directional derivatives. That is, for a direction $h \in \mathbb{F}_p^n$ define the derivative of f in direction h as $D_h f(x) = f(x+h) - f(x)$. Then, f is a polynomial of degree at most d iff

$$D_{h_1} \dots D_{h_{d+1}} f \equiv 0 \qquad \forall h_1, \dots, h_{d+1} \in \mathbb{F}_p^n.$$

Nonclassical polynomials extend this definition to a larger class of objects. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the torus. For a function $f : \mathbb{F}_p^n \to \mathbb{T}$, define its directional derivative in direction $h \in \mathbb{F}_p^n$ as before, as $D_h f(x) = f(x+h) - f(x)$. Then, we define f to be a *nonclassical polynomial of degree at most d* if it is annihilated by

any $d+1$ derivatives,

$$D_{h_1} \dots D_{h_{d+1}} f \equiv 0 \qquad \forall h_1, \dots, h_{d+1} \in \mathbb{F}_p^n.$$

While not immediately obvious, the class of nonclassical polynomials contains the classical polynomials. Let $|\cdot| : \mathbb{F}_p \to \{0, \dots, p-1\} \subset \mathbb{Z}$ denote the natural embedding. If $f : \mathbb{F}_p^n \to \mathbb{F}_p$ is a classical polynomial of degree d then $|f(x)|/p \pmod 1$ is a nonclassical polynomial of degree d. It turns out that as long as $d < p$, these capture all the nonclassical polynomials. However, for $d \geq p$ nonclassical polynomials strictly extend classical polynomials of the same degree. For example, the following is a nonclassical polynomial of degree p:

$$f(x) = \frac{\sum |x_i|}{p^2} \ .$$

See Section 3.2 for more details on nonclassical polynomials.

Correlation bounds for polynomials. We first consider the problem of constructing explicit boolean functions which cannot be approximated by low-degree polynomials. For simplicity, we focus on polynomials defined over $\mathbb{F}_2$, but note that the results below extend to any constant prime finite field. This problem was studied by Razborov [138] and Smolensky [151] in the context of proving lower bounds for $\mathrm{AC}^0(\oplus)$ circuits (and more generally, bounded depth circuits with modular gates modulo a fixed prime). Consider for example the function $\mathrm{MOD}_3 : \{0,1\}^n \to \{0,1\}$, which outputs 1 if the sum of the bits is zero modulo 3, and outputs 0 otherwise. The probability it outputs 0 is $2/3$. They showed that low degree polynomials over $\mathbb{F}_2$ cannot improve this significantly. If $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a polynomial of degree d

then

$$\mathbf{Pr}_{x\in\{0,1\}^n}\left[f(x) = \mathrm{MOD}_3(x)\right] \leq \frac{2}{3} + O\left(\frac{d}{\sqrt{n}}\right).$$

This is sufficient to prove that the MOD_3 function cannot be computed by sub-exponential $\mathrm{AC}^0(\oplus)$ circuits. However, one would like to prove that it cannot even be slightly approximated. Such a result would be a major step towards constructing pseudorandom generators for $\mathrm{AC}^0(\oplus)$ circuits [132, 130], a well known open problem in circuit complexity. It turns out that the Razborov-Smolensky bound is tight for very large degrees, as there exist polynomials of degree $d = \Omega(\sqrt{n})$ which approximate the MOD_3 function with probability 0.99, say. However, it seems to be far from tight for $d \ll \sqrt{n}$, which suggests that an alternative proof technique may be needed.

Viola and Wigderson [170] proved stronger inapproximability results for degrees $d \ll \log n$. These are better described if one considers the correlation of f with the sum of the bits modulo 3. In the following, let $\omega_3 = \exp(2\pi i/3)$ be a cubic root of unity. They showed that if $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a polynomial of degree d then

$$\mathbb{E}_{x\in\{0,1\}^n}\left[(-1)^{f(x)}\omega_3^{x_1+\ldots+x_n}\right] \leq 2^{-\Omega(n/4^d)}.$$

The technique of [170] proves exponential correlation bounds for constant degrees, but decays quickly and becomes trivial at $d = O(\log n)$. Our first result is that this is because of a good reason. Their technique is based on derivatives, and hence this fact extends to nonclassical polynomials. Moreover, it is tight for nonclassical polynomials. In the following, let $e : \mathbb{T} \to \mathbb{C}^*$ be defined as $e(x) = \exp(2\pi i x)$.

Theorem 3.1.1 (Correlation bounds with modular sums for nonclassical polynomials (informal))**.** *Let $f : \mathbb{F}_2^n \to \mathbb{T}$ be a nonclassical polynomial of degree d. Then*

$$\mathbb{E}_{x\in\{0,1\}^n}\left[e(f(x))\omega_3^{x_1+\ldots+x_n}\right] \leq 2^{-\Omega(n/4^d)}.$$

Moreover, for any $\varepsilon > 0$ there exists a nonclassical polynomial $f : \mathbb{F}_2^n \to \mathbb{T}$ of degree $O(\log(n/\varepsilon))$ such that

$$\mathbb{E}_{x\in\{0,1\}^n}\left[e(f(x))\omega_3^{x_1+\ldots+x_n}\right] \geq 1-\varepsilon.$$

So, the Viola-Wigderson technique is bounded for degrees smaller than $O(\log n)$, because it extends to nonclassical polynomials of that degree, for which it is tight. We note that the modulus 3 in Theorem 3.1.1 can be replaced with any fixed odd modulus.

Another boolean function which was shown by Razborov and Smolensky [138, 151] to be hard for $\mathrm{AC}^0(\oplus)$ circuits is the majority function $\mathrm{Maj} : \mathbb{F}_2^n \to \mathbb{F}_2$. The proof relies on the following key fact. If $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a degree d polynomial then

$$\mathbf{Pr}_{x\in\{0,1\}^n}\left[f(x) = \mathrm{Maj}(x)\right] \leq \frac{1}{2} + O\left(\frac{d}{\sqrt{n}}\right). \tag{3.1}$$

Equivalently, this can be presented as a correlation bound

$$\mathbb{E}_{x\in\{0,1\}^n}\left[(-1)^{f(x)}(-1)^{\mathrm{Maj}(x)}\right] \leq O\left(\frac{d}{\sqrt{n}}\right).$$

This is known to be tight for degree $d = 1$ (as say x_1 has correlation $\Omega(1/\sqrt{n})$ with the majority function) and also for $d = \Omega(\sqrt{n})$, since there exist polynomials of that degree which approximate well the majority function, or any symmetric function

for that matter. However, it is not known if these bounds are tight for degrees $1 \ll d \ll \sqrt{n}$. We study this question for nonclassical polynomials. We show that there are nonclassical polynomials of degree $O(\log n)$ with a constant correlation with the majority function.

Theorem 3.1.2 (Correlation bounds with majority for nonclassical polynomials (informal))**.** *There exists a nonclassical polynomial $f : \mathbb{F}_2^n \to \mathbb{T}$ of degree $O(\log n)$ such that*

$$\left|\mathbb{E}\left[e(f(x))(-1)^{\mathrm{Maj}(x)}\right]\right| \geq \Omega(1).$$

So, the Razborov-Smolensky technique separates classical from nonclassical polynomials, since classical polynomials of degree $O(\log n)$ have negligible correlation with the majority function, while as we show above, this is false for nonclassical polynomials.

Exact computation by polynomials. A related problem to correlation bounds is that of exact computation with good probability. For classical polynomials the two problems are equivalent, but this is not the case for nonclassical polynomials. Given a nonclassical polynomial $f : \mathbb{F}_2^n \to \mathbb{T}$, we can ask what is the probability that f is equal to a boolean function, say the majority function. To do so, we identify naturally $\mathbb{F}_2$ with $\{0, 1/2\} \subset \mathbb{T}$, and consider $\mathrm{Maj} : \mathbb{F}_2^n \to \{0, 1/2\}$. We show the following result, which gives a partial answer to the question.

Theorem 3.1.3 (Exact computation of majority by nonclassical polynomials (informal))**.** *Let $f : \mathbb{F}_2^n \to \mathbb{T}$ be a nonclassical polynomial of degree d. Then,*

$$\mathbf{Pr}_{x \in \{0,1\}^n}[f(x) = \mathrm{Maj}(x)] \leq \frac{1}{2} + O\left(\frac{d2^d}{\sqrt{n}}\right).$$

We believe that the bound is not tight, and that, unlike for correlation bounds, nonclassical polynomials should not be able to exactly compute boolean functions better than classical polynomials. Specifically, we ask the following problem.

Open Problem 3.1.4. *Let $f : \mathbb{F}_2^n \to \mathbb{T}$ be a nonclassical polynomial of degree d. Show that*

$$\mathbf{Pr}_{x\in\{0,1\}^n}[f(x) = \mathrm{Maj}(x)] \leq \frac{1}{2} + O\left(\frac{d}{\sqrt{n}}\right).$$

Recently, Bhrushundi, Harsha and Srinivasan [37] resolved the above conjecture positively.

Weak representation of the OR function. We next move to the problem of weak representation of the OR function. Let $p_1, \ldots, p_r$ be distinct primes and let $m = p_1 \ldots p_r$. The goal is to construct a low degree polynomial $f \in \mathbb{Z}_m[x_1, \ldots, x_n]$ such that $f(0^n) = 0$ but $f(x) \neq 0$ for all nonzero $x \in \{0,1\}^n$. Such polynomials stand at the core of locally decodable codes [179, 58, 56, 32, 57], and were further investigated in [151, 11, 15, 14, 13, 12]. There are currently exponential gaps between the best constructions and lower bounds. Barrington, Beigel and Rudich [13] showed that there exist polynomials of degree $O(n^{1/r})$ that weakly represent the OR function. The best lower bound is $\Omega(\log^{1/(r-1)} n)$, due to Barrington and Tardos [12].

The definition of weak representation can be equivalently defined (via the Chinese Remainder Theorem) as follows. There exist polynomials $f_i : \mathbb{F}_{p_i}^n \to \mathbb{F}_{p_i}$ for $i = 1, \ldots, r$ such that $f_1(0^n) = \ldots = f_r(0^n) = 0$ but for any nonzero $x \in \{0,1\}^n$,

there exists an i for which $f_i(x) \neq 0$. This definition can be naturally extended to nonclassical polynomials, where we consider $f_i : \mathbb{F}_{p_i}^n \to \mathbb{T}$. We show that the Barrington-Tardos lower bound extends to nonclassical polynomials, and it is tight up to polynomial factors.

Theorem 3.1.5 (Weak representation of OR for nonclassical polynomials (informal))**.** *Let $p_1, \ldots, p_r$ be distinct primes, and $f_i : \mathbb{F}_{p_i}^n \to \mathbb{T}$ be nonclassical polynomials which weakly represent the OR function. Then*

$$\max \deg(f_i) \geq \Omega(\log^{1/r} n).$$

Moreover, for any fixed prime p, there exists a nonclassical polynomial $f : \mathbb{F}_p^n \to \mathbb{T}$ of degree $O(\log n)$ which weakly represents the OR function.

Thus, the proof technique of Barrington-Tardos cannot extend beyond degree $O(\log n)$, as it applies to nonclassical polynomials as well, for which the $O(\log n)$ bound holds even for prime modulus. We note that unlike in the case of Theorem 3.1.1, where the lower bound proof of [170] extended naturally to nonclassical polynomials, extending the lower bound technique of [12] to nonclassical polynomials requires several nontrivial modifications of the original proof.

Pseudorandom generators for low degree polynomials. Consider for simplicity polynomials over $\mathbb{F}_2$. A distribution D over $\mathbb{F}_2^n$ is said to fool polynomials of degree d with error ε, if for any polynomial $f : \mathbb{F}_2^n \to \mathbb{F}_2$ of degree at most d, we have

$$\left|\mathbf{Pr}_{x \sim D}[f(x) = 0] - \mathbf{Pr}_{x \in \mathbb{F}_2^n}[f(x) = 0]\right| \leq \varepsilon.$$

Distributions which fool linear functions (e.g. $d = 1$) are called small bias generators, and optimal constructions of them (up to polynomial factors) were given in [127, 3], with seed length $O(\log n/\varepsilon)$. A sequence of works [39, 120, 169] showed that small bias generators can be combined to yield generators for larger degree polynomials. The best construction to date is by Viola [169], who showed that the sum of d independent small bias generators with error approximately ε^{2^d} fools degree d polynomials with error ε. Thus, his construction has seed length $O(2^d \log(1/\varepsilon) + d \log n)$, and becomes trivial for $d = \Omega(\log n)$. It is not clear whether it is necessary to require the small bias generators to have smaller error than the required error for the degree d polynomials, and this is the main source for the loss in parameters when considering large degrees.

There is a natural extension of these definitions to nonclassical polynomials. If $f : \mathbb{F}_2^n \to \mathbb{T}$ is a nonclassical polynomial of degree d, then we require that

$$\left|\mathbb{E}_{x\sim D}[e(f(x))] - \mathbb{E}_{x\in\mathbb{F}_2^n}[e(f(x))]\right| \leq \varepsilon.$$

The proof technique of Viola is based on derivatives, and we note here (without proof) that it extends to nonclassical polynomials in a straightforward way. We suspect that it is tight for nonclassical polynomials, however we were unable to show that. Thus, we raise the following open problem.

Open Problem 3.1.6. *Fix $\varepsilon > 0, d \geq 1$. Does there exist a small bias generator with error $\gg \varepsilon^{2^d}$, such that the sum of d independent copies of the generator does not fool degree d nonclassical polynomials with error ε?*

3.1.1 Organisation

We start with some preliminaries in Section 3.2. In Section 3.3, we prove the bounds on approximation of modular sums by nonclassical polynomials. Next, in Section 3.4, we analyze the approximation of the majority function by nonclassical polynomials in the correlation model and the exact computation model. We prove the results on the weak representation of the OR function in Section 3.5.

3.2 Preliminaries

Let $\mathbb{N} = \{1, 2, \ldots\}$ denote the set of positive integers. For $n \in \mathbb{N}$, let $[n] := \{1, 2, \ldots, n\}$. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denote the torus. This is an abelian group under addition. Let $e : \mathbb{T} \to \mathbb{C}^*$ be defined by $e(x) = \exp(2\pi i x)$.

Nonclassical polynomials. Let $\mathbb{F}_p$ be a prime finite field. Given a function $f : \mathbb{F}_p^n \to \mathbb{T}$, its directional derivative in direction $h \in \mathbb{F}_p^n$ is $D_h f : \mathbb{F}_p^n \to T$, given by

$$D_h f(x) = f(x + h) - f(x).$$

Polynomials are defined as functions which are annihilated by repeated derivatives.

Definition 3.2.1 (Nonclassical polynomials)**.** *A function $f : \mathbb{F}_p^n \to \mathbb{T}$ is a polynomial of degree at most d if $D_{h_1} \ldots D_{h_{d+1}} f \equiv 0$ for any $h_1, \ldots, h_{d+1} \in \mathbb{F}_p^n$. The degree of f is the minimal d for which this holds.*

Classic polynomials satisfy this definition. Let $|\cdot|$ denote the natural map from $\mathbb{F}_p$ to $\{0, 1, \ldots, p-1\} \subseteq \mathbb{Z}$. If $P : \mathbb{F}_p^n \to \mathbb{F}_p$ is a (standard) polynomial of degree d, then $f(x) = |P(x)|/p \pmod 1$ is a nonclassical polynomial of degree

d. For degrees $d \leq p$, it turns out that these are the only possible polynomials. However, when $d > p$, there are more polynomials than just these arising from the classical ones, from which the term *nonclassical polynomials* arise. A complete characterization of nonclassical polynomials was developed by Tao and Ziegler [159]. They showed that a function $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leq d$ if and only if it has the following form:

$$f(x_1, \ldots, x_n) = \alpha + \sum_{0 \leq e_1, \ldots, e_n \leq p-1, k \geq 0 : \sum e_i + (p-1)k \leq d} \frac{c_{e_1,\ldots,e_n,k} |x_1|^{e_1} \ldots |x_n|^{e_n}}{p^{k+1}} \pmod 1.$$

Here, $\alpha \in \mathbb{T}$ and $c_{e_1,\ldots,e_n,k} \in \{0, 1, \ldots, p-1\}$ are uniquely determined. The coefficient α is called the *shift* of f, and the largest k for which $c_{e_1,\ldots,e_n,k} \neq 0$ for some $e_1, \ldots, e_n$ is called the *depth* of f. Classical polynomials correspond to polynomials with 0 shift and 0 depth. In this work, we assume without loss of generality that all polynomials have shift 0. Define $\mathbb{U}_{p,k} := \frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$ which is a subgroup of $\mathbb{T}$. Then, the image of polynomials of depth $k-1$ lie in $\mathbb{U}_{p,k}$. We prove the following lemma which shows that nonclassical polynomials can be "translated" to classical polynomials of a somewhat higher degree, at least if we restrict our attention to boolean inputs.

Lemma 3.2.2. *Let $f : \mathbb{F}_p^n \to \mathbb{T}$ be a polynomial of degree d and depth $\leq k-1$. Let $\varphi : \mathbb{U}_{p,k} \to \mathbb{F}_p$ be any function. Then there exists a classical polynomial $g : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree at most $(p^k - 1)d$, such that*

$$g(x) = \varphi(f(x)) \qquad \forall x \in \{0,1\}^n.$$

Proof. By the characterization of nonclassical polynomials, we have

$$f(x) = \sum_{e,j} \frac{c_{e,j} |x_1|^{e_1} \ldots |x_n|^{e_n}}{p^j}$$

where the sum is over $e = (e_1, \ldots, e_n)$ with $e_i \in \{0, \ldots, p-1\}$, $1 \leq j \leq k$ such that $\sum e_i + (p-1)(j-1) \leq d$. We only care about the evaluation of f on the boolean hypercube, which allows for some simplifications. For any $x \in \{0,1\}^n$ we have $|x_1|^{e_1} \ldots |x_n|^{e_n} = \prod_{i \in I} x_i$ where $I = \{i : e_i \neq 0\}$. Thus, we can define an integer polynomial $P(x) = \sum_I c'_I \prod_{i \in I} x_i$ such that

$$f(x) = \frac{P(x)}{p^k} \pmod 1 \qquad \forall x \in \{0,1\}^n,$$

where $c'_I = \sum_{e:\{i:e_i \neq 0\}=I} \sum_j p^{k-j} c_{e,j}$. In particular, note that P has degree at most d. We may further simplify $P(x) = M_1(x) + \ldots + M_t(x)$, where each M_i is a monomial of the form $\prod_{i \in I} x_i$, and monomials may be repeated (indeed, the monomial $\prod_{i \in I} x_i$ is repeated c'_I times). Hence

$$f(x) = \frac{M_1(x) + \ldots + M_t(x)}{p^k} \pmod 1 \qquad \forall x \in \{0,1\}^n.$$

We care about the first k digits in base p of $P(x) = \sum M_i(x)$. These can be captured via the symmetric polynomials, using the fact that $M_i(x) \in \{0,1\}$ for all $x \in \{0,1\}^n$.

The ℓ-th symmetric polynomial in $z = (z_1, \ldots, z_t)$, for $1 \leq \ell \leq t$, is a classical polynomial of degree ℓ defined as

$$S_\ell(z) = \sum_{S \subset [t], |S| = \ell} \prod_{i \in S} z_i.$$

When $z \in \{0,1\}^t$, it follows by Lucas theorem [123] that the i-th digit of $z_1 + \ldots + z_t$ in base p is given by $S_{p^i}(z) \pmod p$.

So, define a polynomial $Q : \mathbb{F}_p^k \to \mathbb{F}_p$ such that $Q(a_0, \ldots, a_{k-1}) = \varphi(\sum a_i p^i / p^k)$ for all $a_0, \ldots, a_{k-1} \in \{0, \ldots, p-1\}$, and polynomials $R_i : \mathbb{F}_p^n \to \mathbb{F}_p$ for

$i = 0, \ldots, k-1$ by $R_i(x) = S_{p^i}(M_1(x), \ldots, M_t(x))$. Note that $\deg(R_i) \leq p^i d$. Define $g(x) = Q(R_0(x), \ldots, R_{k-1}(x))$. Then we have that

$$g(x) = \varphi(f(x)) \qquad \forall x \in \{0,1\}^n.$$

To conclude, we need to bound the degree of g. As monomials in Q raise each variable to degree at most $p-1$, we have $\deg(g) \leq (p-1)\sum \deg(R_i) \leq (p^k - 1)d$. □

Gowers uniformity norms. Let $F : \mathbb{F}^n \to \mathbb{C}$. The (multiplicative) derivative of F in direction $h \in \mathbb{F}^n$ is given by $(\Delta_h F)(x) = F(x+h)\overline{F(x)}$. One can verify that if $f : \mathbb{F}^n \to \mathbb{T}$ and $F = e(f)$ then $\Delta_h F = e(D_h f)$. The d-th Gowers uniformity norm $\|\cdot\|_{U^d}$ is defined as

$$\|F\|_{U^d} := \left(\mathbb{E}_{h_1,\ldots,h_d,x \in \mathbb{F}^n}[\Delta_{h_1} \ldots \Delta_{h_d} F(x)]\right)^{1/2^d}.$$

Observe that $\|F\|_{U^1} = |\mathbb{E}_x[F(x)]|$, which is a semi-norm. For $d \geq 2$, the Gowers uniformity norm turns out to indeed be a norm (but we will not need that). The following lists the properties of the Gowers uniformity norm that we would need. For a proof and further details, see [79].

- Let $f : \mathbb{F}^n \to \mathbb{T}$ and $F = e(f)$. Then $0 \leq \|F\|_{U^d} \leq 1$, where $\|F\|_{U^d} = 1$ if and only if f is a polynomial of degree $\leq d-1$.
- If $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leq d-1$ then $\|Fe(f)\|_{U^d} = \|F\|_{U^d}$ for any $F : \mathbb{F}^n \to \mathbb{C}$.
- If $F(x_1, \ldots, x_n) = F_1(x_1) \ldots F_n(x_n)$ then $\|F\|_{U^d} = \|F_1\|_{U^d} \ldots \|F_n\|_{U_d}$.

- (Gowers-Cauchy-Schwarz) For any $F : \mathbb{F}^n \to \mathbb{C}$ and any $d \geq 1$,

$$0 \leq \|F\|_{U^1} \leq \|F\|_{U^2} \leq \ldots \leq \|F\|_{U^d}.$$

3.3 Approximating modular sums by polynomials

Viola and Wigderson [170] proved that low-degree polynomials over $\mathbb{F}_2$ cannot correlate to the sum modulo m, as long as m is odd. Their proof technique is based on the Gowers uniformity norm. As such, it extends naturally to nonclassical polynomials. We capture that by the following theorem. In the following, let $\omega_m = \exp(2\pi i/m)$ be a primitive m-th root of unity.

Theorem 3.3.1 (Extension of [170] to nonclassical polynomials)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{T}$ be a polynomial of degree $< d$. Let $m \in \mathbb{N}$ be odd. Then for any $a \in \{1, \ldots, m-1\}$,*

$$\mathbb{E}_{x \in \{0,1\}^n} \left[e(f(x)) \cdot \omega_m^{a(x_1 + \ldots + x_n)} \right] \leq \exp(-cn/4^d)$$

where $c = c_m > 0$.

Proof. Let $F(x) = e(f(x)) \cdot \omega_m^{a(x_1+\ldots+x_n)}$. By the properties of the Gowers uniformity norm,

$$|\mathbb{E}_x[F(x)]| \leq \|F\|_{U^d} = \|\omega_m^{a(x_1+\ldots+x_n)}\|_{U^d} = \prod_{i=1}^{n} \|\omega_m^{ax_i}\|_{U^d} = \|e(g)\|_{U^d}^n,$$

where $g : \mathbb{F}_2 \to \mathbb{T}$ is given by $g(0) = 0, g(1) = a/m$. A routine calculation shows that

$$D_{h_1} \ldots D_{h_d} g(x) = \begin{cases} a'/m & \text{if } h_1 = \ldots = h_d = 1, x = 0 \\ -a'/m & \text{if } h_1 = \ldots = h_d = 1, x = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $a' = a2^{d-1}$ is nonzero modulo m. Hence $\|e(g)\|_{U^d}^{2^d} = (1 - 2^{-d}) + 2^{-d}\cos(2\pi a'/m) \leq 1 - 2^{-d} \cdot \Omega(1/m^2)$ and

$$|\mathbb{E}[F]| \leq \left(1 - 2^{-d} \cdot \Omega(1/m^2)\right)^{n/2^d} \leq \exp(-cn/4^d)$$

where $c = \Omega(1/m^2)$. □

This proof technique gives trivial bounds for $d \gg \log n$. Here, we show that this is for a good reason, as there are nonclassical polynomials of degree $O(\log n)$ which well approximate the sum modulo m.

Theorem 3.3.2. *Let $m \in \mathbb{N}$ be odd and fix $a \in \{1, \ldots, m-1\}$. For any $\varepsilon > 0$ there exists a polynomial $f : \mathbb{F}_2^n \to \mathbb{T}$ of degree $\log\left(\frac{n+m}{\varepsilon}\right) + O(1)$ such that*

$$\mathbb{E}_{x \in \{0,1\}^n}\left[e(f(x)) \cdot \omega_m^{a(x_1+\ldots+x_n)}\right] = 1 + u$$

where $|u| \leq \varepsilon$.

Proof. Let $k \geq 1$ to be specific later. Let $r \in \{0, \ldots, m-1\}$ be such that $r \equiv a2^k \pmod m$ and let $A = \frac{r - a2^k}{m} \in \mathbb{Z}$. Define $f : \mathbb{F}_2^n \to \mathbb{T}$ as

$$f(x) = \frac{A(|x_1| + \ldots + |x_n|)}{2^k} \pmod 1.$$

Note that f is a polynomial of degree $\leq k$. For $x \in \{0,1\}^n$, if $x_1 + \ldots + x_n = pm + q$ where $q \in \{0, \ldots, m-1\}$, then

$$f(x) \equiv \frac{A(pm+q)}{2^k} \equiv \frac{rp + \frac{rq}{m}}{2^k} - \frac{aq}{m} = -\frac{aq}{m} + \theta_x \pmod 1,$$

where $0 \leq \theta_x \leq (n+m)/2^k$. We choose $k \geq \log\left(\frac{n+m}{\varepsilon}\right) + c$ for some universal constant c so that $|e(\theta_x) - 1| \leq \varepsilon$ for all x. Hence

$$\left|\mathbb{E}\left[e(f(x)) \cdot \omega_m^{a(x_1+\ldots+x_n)}\right] - 1\right| = \left|\mathbb{E}\left[e(\theta_x) - 1\right]\right| \leq \mathbb{E}\left[|e(\theta_x) - 1|\right] \leq \varepsilon.$$

□

3.4 Approximating majority by nonclassical polynomials

The majority function $\mathrm{Maj} : \mathbb{F}_2^n \to \mathbb{F}_2$ is defined as

$$\mathrm{Maj}(x) = \begin{cases} 0 & \text{if } \sum_{i=1}^n |x_i| \leq n/2 \\ 1 & \text{otherwise} \end{cases}$$

We first show that is correlates well with a nonclassical polynomial of degree $O(\log n)$.

Theorem 3.4.1. *There is a nonclassical polynomial $f : \mathbb{F}_2^n \to \mathbb{T}$ of degree $\log n + 1$ such that*

$$\left|\mathbb{E}\left[(-1)^{\mathrm{Maj}(x)} e(f(x))\right]\right| \geq c,$$

where $c > 0$ is an absolute constant.

Proof. We assume n even for the proof. The proof is similar for odd n. Let $A = \lfloor a\sqrt{n} \rfloor$ for $a > 0$ to be specified later. Let k be the smallest integer such that $2^k \geq n$. Set

$$f(x) = \frac{A(\sum_{i=1}^n |x_i| - n/2)}{2^k}.$$

Note that $\deg(f) = \log n + 1$. Now,

$$\begin{aligned}
&\mathbb{E}\left[(-1)^{\mathrm{Maj}(x)} e(f(x))\right] \\
&= 2^{-n} \sum_{i=0}^{n/2} \binom{n}{i} e\left(A(i-n/2)/2^k\right) - 2^{-n} \sum_{i=n/2+1}^{n} \binom{n}{i} e\left(A(i-n/2)/2^k\right) \\
&= 2^{-n} \sum_{j=1}^{n/2} \binom{n}{n/2-j} e\left(-Aj/2^k\right) - 2^{-n} \sum_{j=1}^{n/2} \binom{n}{n/2-j} e\left(Aj/2^k\right) + 2^{-n} \binom{n}{n/2} \\
&= -2i \cdot 2^{-n} \sum_{j=1}^{n/2} \binom{n}{n/2-j} \sin\left(2\pi Aj/2^k\right) + 2^{-n} \binom{n}{n/2},
\end{aligned}$$

where in the last equation $i = \sqrt{-1}$. Let $C = 2^{-n} \sum_{j=1}^{n/2} \binom{n}{n/2-j} \sin\left(2\pi Aj/2^k\right)$, so that $\left|\mathbb{E}\left[(-1)^{\mathrm{Maj}(x)} e(f(x))\right]\right| \geq 2C$. We will show that $C \geq \Omega(1)$. Let $b > 0$ be a constant to be specified later. We bound

$$C \geq 2^{-n} \sum_{j=1}^{b\sqrt{n}} \binom{n}{n/2-j} \sin\left(2\pi Aj/2^k\right) - \exp(-2b^2),$$

where the error term follows from the Chernoff bound. We set $a = 1/8b$. For all $1 \leq j \leq b\sqrt{n}$ we have $2\pi Aj/2^k \leq \pi/4$. Applying the estimate $\sin(x) \geq x/2$ which holds for all $0 \leq x \leq \pi/4$, we obtain that

$$C \geq \frac{\pi}{32b\sqrt{n}} \cdot 2^{-n} \sum_{j=1}^{b\sqrt{n}} \binom{n}{n/2-j} j - \exp(-2b^2).$$

Now, if b is a large enough constant, standard bounds on the binomial coefficients give that

$$2^{-n} \sum_{j=1}^{b\sqrt{n}} \binom{n}{n/2-j} j = \Omega(\sqrt{n}).$$

Hence, we obtain that

$$C \geq \Omega(1/b) - \exp(-2b^2).$$

If b is chosen a large enough constant, this shows that $C \geq \Omega(1)$ as claimed. $\square$

We next show that the Razborov-Smolensky technique generalizes to nonclassical polynomials when we require the polynomial to exactly compute Maj. Recall that we identify $\mathbb{F}_2$ with $\{0, 1/2\} \subset \mathbb{T}$ and consider $\mathrm{Maj} : \mathbb{F}_2^n \to \{0, 1/2\}$.

Theorem 3.4.2. *Let $f : \mathbb{F}_2^n \to \mathbb{T}$ be a nonclassical polynomial of degree d and depth $< k$. Then,*

$$\mathbf{Pr}_{x \in \{0,1\}^n}[f(x) = \mathrm{Maj}(x)] \leq \frac{1}{2} + O\left(\frac{2^k d}{\sqrt{n}}\right).$$

Proof. Let $\varphi : \mathbb{U}_{2,k} \to \mathbb{F}_2$ be defined as $\varphi(0) = 0$, $\varphi(1/2) = 1$ and choose arbitrarily $\varphi(x)$ for $x \in \mathbb{U}_{2,k} \setminus \{0, 1/2\}$. Applying Lemma 3.2.2, there exists a classical polynomial $g : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $g(x) = \varphi(f(x))$ for all $x \in \mathbb{F}_2^n$, where $\deg(g) \leq (2^k - 1)d$. In particular,

$$\mathbf{Pr}_{x \in \mathbb{F}_2^n}[g(x) = \mathrm{Maj}(x)] \geq \mathbf{Pr}_{x \in \mathbb{F}_2^n}[f(x) = \mathrm{Maj}(x)].$$

Hence, we can apply the Razborov-Smolensky [138, 151] bound to g and conclude that

$$\mathbf{Pr}[f(x) = \mathrm{Maj}(x)] \leq \frac{1}{2} + O\left(\frac{\deg(g)}{\sqrt{n}}\right).$$

$\square$

3.5 Weak representation of the OR function

A set of classical polynomials $f_i : \mathbb{F}_{p_i}^n \to \mathbb{F}_{p_i}$ is said to weakly represent the OR function if they all map 0^n to zero, and for any other point in the boolean hypercube, at least one of them map it to a nonzero value. This definition extends naturally to nonclassical polynomials.

Definition 3.5.1. *Let $p_1, \ldots, p_r$ be distinct primes. A set of polynomials $f_i : \mathbb{F}_{p_i}^n \to \mathbb{T}$ weakly represent the OR function if*

- $f_1(0^n) = \ldots = f_r(0^n) = 0$.
- *For any $x \in \{0,1\}^n \setminus 0^n$, there exists some i such that $f_i(x) \neq 0$.*

It is well known that a single classical polynomial $f : \mathbb{F}_p^n \to \mathbb{F}_p$ which weakly represents the OR function, must have degree at least $n/(p-1)$. This is since $f(x)^{p-1}$ computes the OR function on $\{0,1\}^n$, and hence its multi-linearization (obtained by replacing any power $x_i^{e_i}$, $e_i \geq 1$ with x_i) must be the unique multi-linear extension of the OR function, which has degree n.

We first show that there is a nonclassical polynomial of degree $O(\log n)$ which weakly represents the OR function.

Lemma 3.5.2. *There exists a polynomial $f : \mathbb{F}_p^n \to \mathbb{T}$ of degree $O(p\lceil \log_p n \rceil)$ which weakly represents the OR function.*

Proof. Let $k \geq 1$ be minimal such that $p^k > n$. Define $f(x) = \frac{|x_1| + \ldots + |x_n|}{p^k}$. This is a polynomial of degree $1 + (p-1)(k-1)$. Clearly $f(0^n) = 0$ and $f(x) \neq 0$ for any $x \in \{0,1\}^n \setminus 0^n$. □

We show that allowing for multiple nonclassical polynomials can only improve this simple construction by a polynomial factor.

Theorem 3.5.3. *Let $p_1, \ldots, p_r$ be distinct primes, and let $p = \max(p_1, \ldots, p_r)$. Let $f_i : \mathbb{F}_{p_i}^n \to \mathbb{T}$ be polynomials which weakly represent the OR function. Then at least one of the polynomials must have degree $\Omega((\log_p n)^{1/r})$.*

The proof is an adaptation of the result of Barrington and Tardos [12], who proved similar lower bounds for classical polynomials. We start by showing that a low degree polynomial f with $f(0) = 0$ must have another point x with $f(x) = 0$.

Claim 3.5.4. *Let $f : \mathbb{F}_p^n \to \mathbb{T}$ be a polynomial of degree d and depth $\leq k-1$ such that $f(0) = 0$. If $n > (p^k - 1)d$ then there exists $x \in \{0,1\}^n \setminus 0^n$ such that $f(x) = 0$.*

We note that the bound on n is fairly tight, as $f(x) = (x_1 + \ldots + x_n)/p^k \pmod 1$ violates the conclusion of the claim whenever $n < p^k$.

Proof. Let $\varphi : \mathbb{U}_{p,k} \to \mathbb{F}_p$ be given by $\varphi(0) = 0$, $\varphi(x) = 1$ for all $x \neq 0$. Applying Lemma 3.2.2, there exists a classical polynomial $g : \mathbb{F}_p^n \to \mathbb{F}_p$ of degree $\leq (p^k - 1)d$ such that $g(x) = 0$ if $f(x) = 0$, and $g(x) = 1$ if $f(x) \neq 0$, for all $x \in \{0,1\}^n$. If $f(0^n) = 0$ but $f(x) \neq 0$ for all nonzero $x \in \{0,1\}^n$, then g computes the OR function over $\{0,1\}^n$. Hence, $\deg(g) \geq n$, which leads to a contradiction whenever $n > (p^k - 1)d$. □

We next extend Claim 3.5.4 to a find a common root for a number of polynomials.

Claim 3.5.5. *Let $f_1, \ldots f_r : \mathbb{F}_p^n \to \mathbb{T}$ be polynomials of degree d and depth $\leq k-1$ such that $f_i(0) = 0$ for all $i \in [r]$. If $n > (p^k - 1)dr$ then there exists $x \in \{0,1\}^n \setminus 0^n$ such that $f_i(x) = 0$ for all $i \in [r]$.*

Proof. We construct an interpolating polynomial for $f_1, \ldots, f_r$. Following the proof of Claim 3.5.4, for each f_i there exists a classical polynomial $g_i : \mathbb{F}_p^n \to \mathbb{F}_p$ satisfying

the following. For any $x \in \{0,1\}^n$, if $f_i(x) = 0$ then $g_i(x) = 0$, and if $f_i(x) \neq 0$ then $g_i(x) = 1$. Moreover, $\deg(g_i) \leq (p^k - 1)d$. Define $g : \mathbb{F}_p^n \to \mathbb{F}_p$ as

$$g(x) = 1 - \prod_{i=1}^{r}(1 - g_i(x)).$$

Note that $\deg(g) \leq \sum \deg(g_i) \leq (p^k - 1)dr$. Suppose for contradiction that for every $x \in \{0,1\}^n \setminus 0^n$ there is an $i \in [r]$ such that $f_i(x) \neq 0$. Then $g(0) = 0$ as $f_i(0) = 0$ for all $i \in [r]$, but $g(x) = 1$ for all $x \in \{0,1\}^n \setminus 0^n$. Then g computes the OR function over $\{0,1\}^n$, and hence $\deg(g) \geq n$. This leads to a contradiction whenever $n > (p^k - 1)dr$. □

Next, we argue that the hamming ball of radius d is an interpolating set for polynomials of degree d over $\{0,1\}^n$. In the following, let $B(n,d) = \{x \in \{0,1\}^n : \sum x_i \leq d\}$.

Claim 3.5.6. *Let $f : \mathbb{F}_p^n \to \mathbb{T}$ be a polynomial of degree d such that $f(x) = 0$ for all $x \in B(n,d)$. Then $f(x) = 0$ for all $x \in \{0,1\}^n$.*

Proof. Towards contradiction, let $x^* \in \{0,1\}^n$ be a point such that $f(x^*) \neq 0$, with a minimal hamming weight. By assumption, the hamming weight of x^* is at least $d + 1$. Let $i_1, \ldots, i_{d+1} \in [n]$ be distinct coordinates such that $x^*_{i_1} = \ldots = x^*_{i_{d+1}} = 1$. Let $e_j \in \{0,1\}^n$ be the j-th unit vector, defined as $(e_j)_j = 1$ and $(e_j)_{j'} = 0$ for $j' \neq j$. Define vectors $h_1, \ldots, h_{d+1} \in \mathbb{F}_p^n$ by $h_j = -e_{i_j}$. Since f is a degree d polynomial, we have

$$D_{h_1} \ldots D_{h_{d+1}} f \equiv 0.$$

Evaluating this on x^* gives

$$\sum_{I \subset \{i_1,\ldots,i_{d+1}\}} (-1)^{|I|} f(x^* - \sum_{i \in I} e_i) = 0.$$

However, as we chose x^* with minimal hamming weight such that $f(x^*) \neq 0$, we have $f(x^* - \sum_{i \in I} e_i) = 0$ for all nonempty I. Hence also $f(x^*) = 0$. □

Next, we prove that low degree polynomials must be zero on a large combinatorial box. In the following, we identify subsets $S \subset [n]$ with their indicator in $\{0,1\}^n$.

Lemma 3.5.7. *Let $f : \mathbb{F}_p^n \to \mathbb{T}$ be a polynomial of degree d and depth $\leq k-1$ such that $f(0) = 0$. For $\ell \geq 1$, if $n \geq 2dp^k \ell^{d+1}$ then there exist pairwise disjoint and nonempty sets of variables $S_1, \ldots, S_\ell \subset [n]$ such that*

$$f\left(\sum_{i=1}^{\ell} y_i S_i\right) = 0 \qquad \forall y \in \{0,1\}^\ell.$$

Proof. Fix $a_1, \ldots, a_\ell$ to be determined later such that $n \geq a_1 + \ldots + a_\ell$. Let $A_1, \ldots, A_\ell \subset [n]$ be disjoint subsets of variables of size $|A_i| = a_i$. We will find subsets $S_i \subset A_i$ such that $f(\sum y_i S_i) = 0$ for all $y \in \{0,1\}^\ell$. As we may set the variables outside $A_1, \ldots, A_\ell$ to zero, we assume from now on that $n = a_1 + \ldots + a_\ell$.

First, set $a_1 = p^k d$. Consider the restriction of f to A_1 by setting the remaining variables to zero. By Claim 3.5.4, there exists a nonempty set $S_1 \subset A_1$ such that $f(S_1) = 0$.

Next, suppose that we already constructed $S_1 \subset A_1, \ldots, S_j \subset A_j$ for some $1 \leq j < \ell$, such that $f(\sum y_i S_i) = 0$ for all $y \in \{0,1\}^j$. For each $y \in \{0,1\}^j$, define

a polynomial $f_y : \mathbb{F}_p^{A_{j+1}} \to \mathbb{T}$ by

$$f_y(x') = f\left(\sum_{i=1}^{j} y_i S_i + x'\right)$$

where $x' \in \mathbb{F}_p^{A_{j+1}}$ denotes the variables in A_{j+1}. We will find a common nonzero root for $f_y(x')$.

First, consider only $y \in B(j,d)$. The number of such polynomials is $r = \binom{j}{\leq d} = \sum_{i=0}^{d} \binom{j}{i}$. Applying claim 3.5.5, we have that if we choose $a_{j+1} \geq drp^k$ then there exists $S_{j+1} \subset A_{j+1}$ such that

$$f_y\left(S_{j+1}\right) = 0 \qquad \forall y \in B(j,d).$$

We claim that this implies that $f_y(S_{j+1}) = 0$ for all $y \in \{0,1\}^j$. To see that, define $g : \mathbb{F}_p^j \to \mathbb{T}$ by

$$g(y) = f\left(\sum_{i=1}^{j} y_i S_i + S_{j+1}\right).$$

This a polynomial of degree d, and by Claim 3.5.6, if it is zero for all $y \in B(j,d)$, then it is the zero on all $\{0,1\}^d$. Hence, we have that $f(\sum_{i=1}^{j+1} y_i S_i) = 0$ for all $y \in \{0,1\}^{j+1}$.

We now calculate the parameters. We have $\binom{j}{\leq d} \leq 2j^d$, and hence it suffices to take $a_{j+1} = 2dj^d p^k$. Hence, we need $n \geq n_0$ for

$$n_0 = \sum_{j=1}^{\ell} a_j \leq 2dp^k \sum_{j=1}^{\ell} j^d \leq 2dp^k \ell^{d+1}.$$

□

We are now ready to prove Theorem 3.5.3.

Proof of Theorem 3.5.3. Let $p_1, \ldots, p_r$ be distinct primes, and let $p = \max(p_1, \ldots, p_r)$. Let $f_i : \mathbb{F}_{p_i}^n \to \mathbb{T}$ be polynomials of degree at most d and depth at most $k-1$ which weakly represent the OR function. We fix integers $n \geq \ell_0 = n_0 \geq \ell_1 \ldots \geq \ell_{r-1} \geq \ell_r = 1$ which will be specified later. Applying Lemma 3.5.7 to f_1 with parameter ℓ_1, we get that as long as n is large enough, we can find disjoint nonempty subsets $S_{1,1}, \ldots, S_{1,\ell_1} \subset [n]$ such that $f_1(\sum y_i S_{1,i}) = 0$ for all $y \in \{0,1\}^{\ell_1}$.

Next, consider the restriction of f_2 to the combinatorial cube formed by $\{S_{1,i}\}$. That is, define $f_2' : \mathbb{F}_p^{\ell_1} \to \mathbb{T}$ by $f_2'(y) = f_2(\sum y_i S_{1,i})$. Note that f_2' is a polynomial of degree at most d and depth at most $k-1$. Applying Lemma 3.5.7 to f_2' with parameter ℓ_2, we get that as long as ℓ_1 is large enough, we can find disjoint nonempty subsets $S'_{2,1}, \ldots, S'_{2,\ell_2} \subset [\ell_1]$ such that $f_2'(\sum y_i S'_{2,i}) = 0$ for all $y \in \{0,1\}^{\ell_2}$. Define $S_{2,1}, \ldots, S_{2,\ell_2} \subset [n]$ by $S_{2,i} = \cup_{j \in S'_{2,i}} S_{1,j}$. Then $S_{2,1}, \ldots, S_{2,\ell_2}$ are disjoint nonempty subsets of $[n]$, such that

$$f_1\left(\sum_{i=1}^{\ell_2} y_i S_{2,i}\right) = f_2\left(\sum_{i=1}^{\ell_2} y_i S_{2,i}\right) = 0 \qquad \forall y \in \{0,1\}^{\ell_2}.$$

Continuing in this fashion, we ultimately find disjoint nonempty subsets $S_{r,1}, \ldots, S_{r,\ell_r} \subset [n]$ such that

$$f_1\left(\sum_{i=1}^{\ell_r} y_i S_{r,i}\right) = \ldots = f_r\left(\sum_{i=1}^{\ell_r} y_i S_{r,i}\right) = 0 \qquad \forall y \in \{0,1\}^{\ell_r}.$$

In particular, $f_1, \ldots, f_r$ cannot weakly represent the OR function. This argument requires that for each $0 \leq i \leq r-1$, $\ell_i \geq 2dp^k \ell_{i+1}^{d+1}$, which can be satisfied if

$$n \geq n_0 = (2dp^k)^{(d+1)^{r-1}}.$$

Now, $k \leq d/(p-1)+1$ and hence $p^k \leq p^{d/(p-1)+1} \leq 2^d p$. As we can trivially bound $2d \leq 2^d$ we obtain the simplified bound

$$n_0 \leq 2^{4(d+1)^r \cdot \log p}.$$

Thus, if $f_1, \ldots, f_r$ do weakly represent the OR function, at least one of the must have degree $d \geq \Omega((\log_p n)^{1/r})$. $\square$

Chapter 4

The List Decoding Radius of RM codes over small prime fields

4.1 Introduction

The concept of *list decoding* was introduced by Elias [60] and Wozencraft [176] to decode *error correcting codes* beyond half the minimum distance. The objective of list decoding is to output all the codewords within a specified radius around the received word. After the seminal results of Goldreich and Levin [67] and Sudan [153] which gave list decoding algorithms for the Hadamard code and the Reed-Solomon code respectively, there has been tremendous progress in designing list decodable codes. See the excellent surveys of Guruswami [89, 88] and Sudan [154].

List decoding has applications in many areas of computer science including hardness amplification in complexity theory [155, 165], derandomization [168], construction of hard core predicates from one way functions [67, 2], construction of extractors and pseudorandom generators [157, 149] and computational learning [116, 101]. Despite so much progress, the largest radius up to which list decoding

is tractable is still a fundamental open problem even for well studied codes like Reed-Solomon (univariate polynomials) and Reed-Muller codes (multivariate polynomials). The goal of this work is to analyse Reed-Muller codes over small fields and small degree.

Reed-Muller codes (RM codes) were discovered by Muller in 1954. Fix a finite field $\mathbb{F} = \mathbb{F}_q$. Let $d \in \mathbb{N}$. The RM code $\mathrm{RM}_{\mathbb{F}}(n, d)$ is defined as follows. The message space consists of degree $\leq d$ polynomials in n variables over $\mathbb{F}$ and the codewords are evaluation of these polynomials on $\mathbb{F}^n$. Let $\delta_p(d)$ denote the normalized distance of $\mathrm{RM}_{\mathbb{F}}(n, d)$. Let $d = a(q-1) + b$ where $0 \leq b < q - 1$. We have

$$\delta_{\mathbb{F}}(d) = \frac{1}{q^a}\left(1 - \frac{b}{q}\right).$$

RM codes are one of the most well studied error correcting codes. Many of the applications in computer science involves low degree polynomials over small fields, namely RM codes. Given a received word $g : \mathbb{F}^n \to \mathbb{F}$ the objective is to output the list of codewords (e.g. low-degree polynomials) that lie within some distance of g. Typically we will be interested in regimes where list size is either independent of n or polynomial in the block length $\mathbb{F}^n$.

4.1.1 Previous Work

Let $\mathcal{P}_d(\mathbb{F}^n)$ denote the class of degree $\leq d$ polynomials $f : \mathbb{F}^n \to \mathbb{F}$. Let dist denote the normalized Hamming distance. For $\mathrm{RM}_{\mathbb{F}}(n, d)$, $\eta > 0$, let

$$\ell_{\mathbb{F}}(n, d, \eta) := \max_{g:\mathbb{F}^n \to \mathbb{F}} \left|\{f \in \mathcal{P}_d(\mathbb{F}^n) : \mathrm{dist}(f, g) \leq \eta\}\right|.$$

Let $\mathrm{LDR}_{\mathbb{F}}(n,d)$ (short for *list decoding radius*) be the maximum η for which $\ell_{\mathbb{F}}(n,d,\eta-\varepsilon)$ is upper bounded by a constant depending only on $\varepsilon, |\mathbb{F}|, d$ for all $\varepsilon > 0$.

It is easy to see that $\mathrm{LDR}_{\mathbb{F}}(n,d) \leq \delta_{\mathbb{F}}(d)$. The difficulty lies in proving a matching lower bound. The first breakthrough result was in the setting of $d = 1$ over $\mathbb{F}_2$ (Hadamard Codes) where Goldreich and Levin showed that $\mathrm{LDR}_{\mathbb{F}_2}(n,1) = \delta_{\mathbb{F}_2}(1) = 1/2$ [67]. Later, Goldreich, Rubinfield and Sudan [68] generalized the field to obtain $\mathrm{LDR}_{\mathbb{F}}(n,1) = \delta_{\mathbb{F}}(1) = 1 - 1/|\mathbb{F}|$. In the setting of $d < |\mathbb{F}|$, Sudan, Trevisan and Vadhan [155] showed that $\mathrm{LDR}_{\mathbb{F}}(n,d) \geq 1 - \sqrt{2d/|\mathbb{F}|}$ improving previous work by Arora and Sudan [5], Goldreich *et al* [68] and Pellikaan and Wu [133]. A crucial result that was a bulding block in the multivariate setting was the problem of list decoding Reed-Solomon codes which was analysed by Sudan [153] and Guruswami and Sudan [90]. The list decoding radius obtained above essentially attains the Johnson radius, which is a radius such that for any code over $\mathbb{F}$ with normalized minimum distance δ, the list decoding radius (LDR) is at least

$$\mathrm{J}_{\mathbb{F}}(\delta) := \left(1 - \frac{1}{|\mathbb{F}|}\right)\left(1 - \sqrt{1 - \frac{|\mathbb{F}|\delta}{|\mathbb{F}| - 1}}\right).$$

There have been few results that show list decodability beyond the Johnson radius [53, 74].

In 2008, Gopalan, Klivans and Zuckerman [74] showed that $\mathrm{LDR}_{\mathbb{F}_2}(n,d) = \delta_{\mathbb{F}_2}(d)$. This beats the Johnson radius already for $d \geq 2$. The list decoding algorithm in [74] is a generalization of the Goldreich-Levin algorithm [67]. However their algorithm crucially depends on the fact that the ratio of minimum distance to unique decoding radius is equal to 2 which is the size of the field. Therefore, it does not

generalize to higher fields (except for some special cases). They pose the following conjecture.

Conjecture 2 ([74]). *For all constants d and all fields $\mathbb{F}$,* $\mathrm{LDR}_{\mathbb{F}}(n,d) = \delta_{\mathbb{F}}(d)$.

An important contribution of [74] is an algorithm for list decoding that outputs the list of codewords up to radius η efficiently assuming $\ell_{\mathbb{F}}(n,d,\eta)$ is bounded.

It was also shown [74] that $\mathrm{LDR}_{\mathbb{F}}(n,d) \geq \frac{1}{2}\delta_{\mathbb{F}}(d-1)$ and this beats the Johnson radius already when d is large. It is believed [74, 73] that the hardest case is the setting of small d. An important step in this direction was taken in [73] that considered quadratic polynomials and showed that $\mathrm{LDR}_{\mathbb{F}}(n,2) = \delta_{\mathbb{F}}(2)$ for all fields $\mathbb{F}$ and thus proved the conjecture for $d = 2$. In the setting of $\mathbb{F}_2$, Kaufman, Lovett and Porat [107] showed tight list sizes for radii beyond the minimum distance.

4.1.2 Our Results

As mentioned before, the algorithmic problem of list decoding was reduced to the combinatorial problem in [74]. Our main theorem is a resolution of Conjecture 3 for prime fields. We note that prior to this, the conjecture was open even in the $d < |\mathbb{F}|$ case.

Theorem 5. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta_{\mathbb{F}}(d) - \varepsilon) \leq c_{p,d,\varepsilon}.$$

Remark 4.1.1 (Algorithmic Implications). *As mentioned above, using the reduction of algorithmic list decoding to combinatorial list decoding in [74] along with*

Theorem 5, for fixed prime fields, d and $\varepsilon > 0$, we now have list decoding algorithms in both the global setting (running time polynomial in $|\mathbb{F}|^n$) and the local setting (running time polynomial in n^d).

Next, we study list sizes for radii which are larger than the minimal radius of the code. We give bounds which capture the correct exponent of n for all radii. This extends the results of Kaufman, Lovett and Porat [107] who studied Reed-Muller codes over $\mathbb{F}_2$, to all prime fields.

Theorem 6. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $\varepsilon > 0$ and $e \leq d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta_{\mathbb{F}}(e) - \varepsilon) \leq \exp\left(c_{p,d,\varepsilon} n^{d-e}\right)$$

Remark 4.1.2. *The exponent of n in Theorem 6 is tight, as the following example shows. Let $e = a(p-1) + b$ with $0 \leq b < p-1$. Consider polynomials of the form*

$$P(x) = \left(\prod_{i=i}^{a} (x_i^{p-1} - 1)\right) \left(\prod_{j=1}^{b} (x_{a+1} - j)\right) (x_{a+2} + Q(x_{a+3}, \ldots, x_n))$$

for all polynomials Q of degree $d-e$. Observe that $\mathbf{Pr}[P(x) \neq 0] = \frac{1}{p^a}\left(1 - \frac{b}{p}\right)\left(1 - \frac{1}{p}\right) = \delta(e)(1 - 1/p)$. The number of such polynomials is $\exp(c' n^{d-e})$ for some $c' = c'_{p,d,e}$.

4.1.3 Proof overview

Previous results have mostly relied on the idea of local correction of the RM code. The work of [73] uses (linear) Fourier analysis which does not seem to go beyond quadratic polynomials. We use tools from higher order Fourier analysis to resolve the conjecture. We think of $\mathbb{F} = \mathbb{F}_p, d, \varepsilon$ as constants. For a received word

$g : \mathbb{F}^n \to \mathbb{F}$ our goal is to upper bound $|\{f \in \mathcal{P}_d(\mathbb{F}^n) : \mathrm{dist}(f, g) \leq \eta\}|$. For simplicity of exposition, we assume in the proof overview that $d < |\mathbb{F}|$. The general case is somewhat more technical, as it requires the introduction of nonclassical polynomials.

A weak regularity (A low complexity proxy for the received word). The first step is an extension of the Frieze-Kannan weak regularity [62] which would allow us to move from an arbitrary received word g to a "low complexity" received word. We note that a somewhat similar idea appeared also in [167].

Let X, Y be finite sets and let $P(Y) := \{f : Y \to \mathbb{R}_{\geq 0} : \sum_{y \in Y} f(y) = 1\}$ be the probability simplex over Y. We view functions $f : X \to P(Y)$ as randomized functions from X to Y. For $f, g : X \to P(Y)$ we define

$$\mathbf{Pr}_x[f(x) = g(x)] := \mathbb{E}_x \langle f(x), g(x) \rangle.$$

Given $\varepsilon > 0$, any function $g : X \to P(Y)$ and a collection F of functions $f : X \to P(Y)$, one can find a collection of $c := 1/\varepsilon^2$ functions $h_1, \ldots, h_c \in F$ and a *proxy* $g_1 : X \to P(Y)$ for g, such that g_1 is determined by $h_1(x), \ldots, h_c(x)$ and such that g_1 is indistinguishable from g with respect to F.

Lemma 4.3.1. *Let $g : X \to P(Y)$, $\varepsilon > 0$, and F be a collection of functions $f : X \to P(Y)$. Then there exist $c \leq 1/\varepsilon^2$ functions $h_1, h_2, \ldots, h_c \in F$ and a function $\Gamma : P(Y)^c \to P(Y)$ such that for all $f \in F$,*

$$|\mathbf{Pr}[g(x) = f(x)] - \mathbf{Pr}[\Gamma(h_1(x), h_2(x), \ldots, h_c(x)) = f(x)]| \leq \varepsilon.$$

In our case, $X = \mathbb{F}^n$, $Y = \mathbb{F}$ and $F = \mathcal{P}_d(\mathbb{F}^n)$. When F is a family of "deterministic" functions $f : X \to Y$, as it is in our case, we can obtain one-sided approximation using only deterministic functions $h_1, \ldots, h_c$.

Corollary 4.3.3. *Let $g : X \to Y$, $\varepsilon > 0$, and F be a collection of functions $f : X \to Y$. Then there exist $c \leq 1/\varepsilon^2$ functions $h_1, h_2, \ldots, h_c \in F$ such that for every $f \in F$, there is a function $\Gamma_f : Y^c \to Y$ such that*

$$\mathbf{Pr}_x[\Gamma_f(h_1(x), \ldots, h_c(x)) = f(x)] \geq \mathbf{Pr}_x[g(x) = f(x)] - \varepsilon.$$

Strong regularity applied to $\mathcal{H}$. The collection of polynomials $\mathcal{H} = \{h_1, \ldots, h_c\} \subset \mathcal{P}_d(\mathbb{F}^n)$ defines a partition of the input space $\mathbb{F}^n$ into *atoms* $\{x \in \mathbb{F}^n : h_1(x) = a_1, \ldots, h_c(x) = a_c\}$. We next regularize $\mathcal{H}$. The objective of regularization is to further refine the partition into smaller atoms with the goal that the polynomials $h_1, \ldots, h_c$ are "pseudo-random". Formally, we require the polynomials to be inapproximable by lower degree polynomials, which is equivalent to having negligible Gowers uniformity norm. This ensures, for example, that for uniformly random X in $\mathbb{F}^n$, the distribution $(h_1(X), \ldots, h_c(X))$ is close to uniform over the atoms. This process of regularization was introduced by [80] and is now standard in higher-order Fourier analysis. Let $\mathcal{H}' = \{h'_1, \ldots, h'_{c'}\} \subset \mathcal{P}_d(\mathbb{F}^n)$ be the regularized $\mathcal{H}$ that satisfies the above properties, where $c' = c'(p, d, c)$.

Structure of polynomials close to low complexity received words. Fix now an $f \in \mathcal{P}_d(\mathbb{F}^n)$ such that $\mathrm{dist}(f, g) \leq \delta_p(d) - \varepsilon$. We will show that f must be

determined by $\mathcal{H}'$. That is,

$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x))$$

for some $F : \mathbb{F}^{c'} \to \mathbb{F}$. This will bound the number of such functions by $p^{p^{c'}}$, which is independent of n.

In order to achieve that, we regularize the family of polynomials $\mathcal{H}' \cup \{f\}$. By choosing regularity parameters appropriately, we can assure that only f decomposes further,

$$f = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x))$$

where $\mathcal{H}'' = \{h_1, \ldots, h'_{c'}, h''_1, \ldots, h''_{c''}\}$ is regular. Moreover, for $G_f(h'_1(x), \ldots, h'_{c'}(x)) = \Gamma_f(h_1(x), \ldots, h_c(x))$, we know that

$$\mathbf{Pr}[f(x) = G_f(h'_1(x), \ldots, h'_{c'}(x))] \geq 1 - \delta_p(d) + \varepsilon/2.$$

The regularity of $\mathcal{H}''$ allows us to reduce the question to that of the structure of F vs G_f. We then show, by a variant of the Schwartz-Zippel lemma, that such an approximation can only exist when F does not depend on $h''_1, \ldots, h''_{c''}$. The bound for larger radii $\delta_{\mathbb{F}}(e) - \varepsilon$ with $e < d$ follows along similar lines. We show that in the decomposition above, since $\mathbf{Pr}[F = G_f] > 1 - \delta_{\mathbb{F}}(e) + \varepsilon/2$, this can only occur when $h''_1, \ldots, h''_{c''}$ have degree at most $d - e$. As the number of such polynomials is exponential in n^{d-e}, we derive similar bounds for the number of functions f.

4.2 Preliminaries

4.2.1 Notation

Let $\mathbb{N}$ denote the set of positive integers. For $n \in \mathbb{N}$, let $[n] := \{1, 2, \ldots, n\}$. We use $y = x \pm \varepsilon$ to denote $y \in [x - \varepsilon, x + \varepsilon]$. Let $\mathbb{T}$ denote the torus $\mathbb{R}/\mathbb{Z}$. This is an abelian group under addition. For $n \in \mathbb{N}$, and $x, y \in \mathbb{C}^n$, let $\langle x, y \rangle := \sum_{i=1}^{n} x_i \overline{y_i}$ where $\overline{a}$ is the conjugate of a. Let $||x||_2 := \sqrt{\langle x, x \rangle}$.

Fix a prime field $\mathbb{F} = \mathbb{F}_p$. Let $|.|$ denote the natural map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\} \in \mathbb{Z}$. Let $e : \mathbb{T} \to \mathbb{C}$ be the map $e(x) := e^{2\pi i x}$. Let $e_p : \mathbb{F} \to \mathbb{C}$ be the map $e_p(x) = e(\frac{|x|}{p})$. For an integer $k \geq 0$, let $\mathbb{U}_k := \frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$. Note that $\mathbb{U}_k$ is a subgroup of $\mathbb{T}$. Let $\iota : \mathbb{F} \to \mathbb{U}_1$ be the bijection $\iota(a) = \frac{|a|}{p} \pmod 1$.

For a finite set X and $n \in \mathbb{N}$, with $f : X \to \mathbb{C}^n$, we write $\mathbb{E}_x f(x)$ to denote $\frac{1}{|X|} \sum_{x \in X} f(x)$. We define $||f||_2 := \sqrt{\mathbb{E}_x ||f(x)||_2^2}$. If $g : X \to \mathbb{C}^n$, we have $\langle f, g \rangle := \mathbb{E}_x \langle f(x), g(x) \rangle$. Let Y be a finite set. Let $P(Y) := \{f : Y \to \mathbb{R}_{\geq 0} : \sum_{y \in Y} f(y) = 1\}$ denote the probability simplex on Y. We shall write randomized functions by mapping them to the simplex. Thus, for $f, g : X \to P(Y)$ we define

$$\mathbf{Pr}_x[f(x) = g(x)] := \mathbb{E}_x \langle f(x), g(x) \rangle.$$

If $f : X \to Y$ is a deterministic function, then we embed Y into $P(Y)$ in the obvious way, and consider $f : X \to P(Y)$ with $f(x)_y = 1$ if $f(x) = y$ when viewed as a function to Y, and $f(x)_{y'} = 0$ for all $y' \in Y \setminus \{y\}$.

4.2.2 Polynomials

Definition 4.2.1 (Derivative)**.** *Given a function $f : \mathbb{F}^n \to \mathbb{T}$ and $a \in \mathbb{F}^n$, define the derivative of f in direction a as $D_a f : \mathbb{F}^n \to \mathbb{T}$ as $D_a f(x) = f(x+a) - f(x)$ for $x \in \mathbb{F}^n$.*

Definition 4.2.2 (Nonclassical Polynomial or Polynomial)**.** *Let $d \in \mathbb{N}$. Then $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leq d$ if for all $a_1, \ldots, a_{d+1}, x \in \mathbb{F}^n$,*

$$\left(D_{a_1} \ldots D_{a_{d+1}} f\right)(x) = 0. \tag{4.1}$$

The degree of f denoted by $\deg(f)$ is the smallest such $d \in \mathbb{N}$ for which the above holds. If the image of f lies in $\mathbb{U}_1$ then f is called a classical polynomial of degree d. When $d < |\mathbb{F}|$, it is known that all the polynomials of degree d satisfying (4.1) are classical polynomials. However, when $d \geq |\mathbb{F}|$, there exist nonclassical polynomials. We write $\mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T})$ to denote the class of degree $\leq d$ polynomials. Unless explicitly specified, a polynomial is a (potentially) nonclassical polynomial. The following lemma from [160] characterizes polynomials.

Lemma 4.2.3 ([160], Lemma 1.7)**.** *Let $d \in \mathbb{N}$.*

- *A function $f : \mathbb{F}^n \to \mathbb{T}$ is a polynomial of degree $\leq d$ if and only if $D_a f$ is a polynomial of degree $\leq d-1$ for all $a \in \mathbb{F}^n$.*
- *A function $f : \mathbb{F}^n \to \mathbb{T}$ is a classical polynomial with $\deg(f) \leq d$ if $f = \iota \circ P$ where $P : \mathbb{F}^n \to \mathbb{F}$ is of the form*

$$P(x_1, \ldots, x_n) = \sum_{0 \leq d_1, \ldots, d_n \leq p-1 : \sum_i d_i \leq d} c_{d_1, \ldots, d_n} \prod_{i=1}^{n} x_i^{d_i},$$

where $c_{d_1,\ldots,d_n} \in \mathbb{F}$ *are unique.*

- *A function* $f : \mathbb{F}^n \to \mathbb{T}$ *is a polynomial with* $deg(f) \leq d$ *if* f *is of the form*

$$f(x_1, \ldots, x_n) = \alpha + \sum_{0 \leq d_1,\ldots,d_n \leq p-1, k \geq 0 : \sum_i d_i \leq d-k(p-1)} \frac{c_{d_1,\ldots,d_n,k} \prod_{i=1}^{n} |x_i|^{d_i}}{p^{k+1}} \pmod 1,$$

where $c_{d_1,\ldots,d_n,k} \in \{0, \ldots, p-1\}$ *and* $\alpha \in \mathbb{T}$ *are unique.* α *is called the shift of* f *and the largest* k *such that some* $c_{d_1,\ldots,d_n,k} \neq 0$ *is the depth of* f*, denoted by* $\mathrm{depth}(f)$*. Note that classical polynomials have* 0 *shift and* 0 *depth.*

- *If* $f : \mathbb{F}^n \to \mathbb{T}$ *is a polynomial with* $\mathrm{depth}(f) = k$*, then its image lies in a coset of* $\mathbb{U}_{k+1}$*.*

- *If* $f : \mathbb{F}^n \to \mathbb{T}$ *is a polynomial such that* $\deg(f) = d$ *and* $\mathrm{depth}(f) = k$*, then* $\deg(pf) = \max(d-p+1, 0)$ *and* $\mathrm{depth}(pf) = k-1$*. Also, if* $c \in \{1, \ldots, p-1\}$ *then the degree and depth of* cf *remain unchanged.*

Throughout the article, we assume without loss of generality that nonclassical polynomials have zero shift.

4.2.3 Rank and Polynomial Factors

Definition 4.2.4 (Rank)**.** *Let* $d \in \mathbb{N}$ *and* $f : \mathbb{F}^n \to \mathbb{T}$*. Then* $\mathrm{rank}_d(f)$ *is defined as the smallest integer* r *such that there exist polynomials* $h_1, \ldots, h_r : \mathbb{F}^n \to \mathbb{T}$ *of degree* $\leq d-1$ *and a function* $\Gamma : \mathbb{T}^r \to \mathbb{T}$ *such that* $f(x) = \Gamma(h_1(x), \ldots, h_r(x))$*. If* $d = 1$*, then the rank is* 0 *if* f *is a constant function and is* ∞ *otherwise. If* f *is a polynomial, then* $\mathrm{rank}(f) = \mathrm{rank}_d(f)$ *where* $d = \deg(f)$*.*

Definition 4.2.5 (Factor)**.** *Let X be a finite set. Then a factor $\mathcal{B}$ is a partition of the set X. The subsets in the partition are called atoms.*

For sets X and Y, and a factor $\mathcal{B}$ of X, a function $f : X \to P(Y)$ is said to be measurable with respect to $\mathcal{B}$ if it is constant on the atoms of $\mathcal{B}$. The average of f over $\mathcal{B}$ is $\mathbb{E}[f|\mathcal{B}] : X \to P(Y)$ defined as

$$\mathbb{E}[f|\mathcal{B}](x) = \mathbb{E}_{y \in \mathcal{B}(x)}[f(y)]$$

where $\mathcal{B}(x)$ is the atom containing x. Clearly, $\mathbb{E}[f|\mathcal{B}]$ is measurable with respect to $\mathcal{B}$.

A collection of functions $h_1, \ldots, h_c : X \to Y$ defines a factor $\mathcal{B}$ whose atoms are $\{x \in X : h_1(x) = y_1, \ldots, h_c(x) = y_c\}$ for every $(y_1, \ldots, y_c) \in Y^c$. We use $\mathcal{B}$ to also denote the map $x \mapsto (h_1(x), \ldots, h_c(x))$. A function f is measurable with respect to a collection of functions if it is measurable with respect to the factor the collection defines.

Definition 4.2.6 (Polynomial Factor)**.** *A polynomial factor $\mathcal{B}$ is a factor defined by a collection of polynomials $\mathcal{H} = \{h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{T}\}$ and the factor is written as $\mathcal{B}_{\mathcal{H}}$. The degree of the factor is the maximum degree of $h \in \mathcal{H}$.*

Let $|\mathcal{B}|$ be the number of polynomials defining the factor. If $\text{depth}(h_i) = k_i$ above, then we define $||\mathcal{B}|| := \prod_{i=1}^{c} p^{k_i+1}$ to be the number of (possibly empty) atoms.

Definition 4.2.7 (Rank and Regularity of Polynomial Factor)**.** *Let $\mathcal{B}$ be a polynomial factor defined by $h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{T}$ such that* $\text{depth}(h_i) = k_i$ *for $i \in [c]$.*

Then, the rank of $\mathcal{B}$ is the least integer r such that there exists $(a_1, \ldots, a_c) \in \mathbb{Z}^c$, $\left(a_1 \mod p^{k_1+1}, \ldots, a_c \mod p^{k_c+1}\right) \neq (0, \ldots, 0)$ for which the linear combination $h(x) := \sum_{i=1}^{c} a_i h_i(x)$ has $\mathrm{rank}_d(h) \leq r$ where $d = \max_i \deg(a_i h_i)$. For a non decreasing function $r : \mathbb{N} \to \mathbb{N}$, a factor $\mathcal{B}$ is r-regular if its rank is at least $r(|\mathcal{B}|)$.

Definition 4.2.8 (Semantic and Syntactic refinement)**.** *Let $\mathcal{B}$ and $\mathcal{B}'$ be polynomial factors on $\mathbb{F}^n$. A factor $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$, denoted by $\mathcal{B}' \succeq_{syn} \mathcal{B}$ if the set of polynomials defining $\mathcal{B}$ is a subset of the set of polynomials defining $\mathcal{B}'$. It is a semantic refinement, denoted by $\mathcal{B}' \succeq_{sem} \mathcal{B}$ if for every $x, y \in \mathbb{F}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$.*

We will use the following regularity lemma proved in [26].

Lemma 4.2.9 (Polynomial Regularity Lemma [26])**.** *Let $r : \mathbb{N} \to \mathbb{N}$ be a non-decreasing function and $d \in \mathbb{N}$. Then there is a function $C_{r,d}^{(6.4.6)} : \mathbb{N} \to \mathbb{N}$ such that the following is true. Let $\mathcal{B}$ be a factor defined by polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ of degree at most d. Then, there is an r-regular factor $\mathcal{B}'$ defined by polynomials $Q_1, \ldots, Q_{c'} : \mathbb{F}^n \to \mathbb{T}$ of degree at most d such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$ and $c' \leq C_{r,d}^{(6.4.6)}(c)$.*

Moreover if $\mathcal{B} \succeq_{sem} \hat{\mathcal{B}}$ for some polynomial factor $\hat{\mathcal{B}}$ that has rank at least $r(c') + c' + 1$, then $\mathcal{B}' \succeq_{syn} \hat{B}$.

The next lemma shows that a regular factor has atoms of roughly equal size.

Lemma 4.2.10 (Size of atoms [26])**.** *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of rank at least $r_d^{(6.4.9)}(\varepsilon)$ defined by polynomials $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ of degree at most*

d such that $\mathrm{depth}(P_i) = k_i$ *for* $i \in [c]$. *For every* $b \in \otimes_{i=1}^{c} \mathbb{U}_{k_i+1}$,

$$\mathbf{Pr}_x[\mathcal{B}(x) = b] = \frac{1}{||\mathcal{B}||} \pm \varepsilon.$$

Finally, we shall need the following lemma which shows that a function of high rank polynomials has the degree one expects.

Lemma 4.2.11 (Preserving degree [26])**.** *Let* $d > 0$ *be an integer and let* $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{T}$ *be polynomials of degree at most* d *that form a factor of* $\mathrm{rank} \geq r_d^{(6.4.16)}(c)$. *Let* $\Gamma : \mathbb{T}^c \to \mathbb{T}$ *be an arbitrary function. Let* $F : \mathbb{F}^n \to \mathbb{T}$ *be defined by* $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$, *and assume that* $\deg(F) = d'$. *Then, for every collection of polynomials* $Q_1, \ldots, Q_c : \mathbb{F}^n \to \mathbb{T}$ *with* $\deg(Q_i) \leq \deg(P_i)$ *and* $\mathrm{depth}(Q_i) \leq \mathrm{depth}(P_i)$, *if* $G : \mathbb{F}^n \to \mathbb{T}$ *is defined by* $G(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$, *then* $\deg(G) \leq d'$.

4.3 Weak Regularity

Let X and Y be finite sets. Recall that $P(Y) := \{f : Y \to \mathbb{R}_{\geq 0} : \sum_{y \in Y} f(y) = 1\}$ is the probability simplex on Y. As mentioned before, we shall write randomized functions by mapping them to the simplex. Thus for $f, g : X \to P(Y)$ we have

$$\mathbf{Pr}_x[f(x) = g(x)] := \mathbb{E}_x \langle f(x), g(x) \rangle.$$

Lemma 4.3.1. *Let* $g : X \to P(Y)$, $\varepsilon > 0$, *and* F *be a collection of functions* $f : X \to P(Y)$. *Then there exist* $c \leq 1/\varepsilon^2$ *functions* $h_1, h_2, \ldots, h_c \in F$ *and a function* $\Gamma : P(Y)^c \to P(Y)$ *such that for all* $f \in F$,

$$|\mathbf{Pr}[g(x) = f(x)] - \mathbf{Pr}[\Gamma(h_1(x), h_2(x), \ldots, h_c(x)) = f(x)]| \leq \varepsilon.$$

Proof. We construct $\mathcal{H} = \{h_1, \ldots, h_c\} \subseteq F$ such that, if $\mathcal{B}_{\mathcal{H}}$ is the factor of X induced by $\mathcal{H}$, then for all $f \in F$

$$|\mathbf{Pr}[\mathbb{E}[g|\mathcal{B}_{\mathcal{H}}] = f(x)] - \mathbf{Pr}[g(x) = f(x)]| \leq \varepsilon.$$

We then set $\Gamma : P(Y)^c \to P(Y)$ so that $\Gamma(h_1(x), \ldots, h_c(x)) = \mathbb{E}[g|\mathcal{B}_{\mathcal{H}}]$. In the following we shorthand $g_{\mathcal{H}} = \mathbb{E}[g|\mathcal{B}_{\mathcal{H}}]$. We consider the following variant of the Frieze-Kannan weak regularity algorithm [62].

- Initialize $\mathcal{H} = \emptyset$
- While there exists $f \in F$ such that $|\mathbf{Pr}[g_{\mathcal{H}}(x) = f(x)] - \mathbf{Pr}[g(x) = f(x)]| > \varepsilon$
  - Update $\mathcal{H} = \mathcal{H} \cup \{f\}$

The lemma follows from the following claim, which shows that we update $\mathcal{H}$ at most $1/\varepsilon^2$ times. Let $\|g_{\mathcal{H}}\|_2^2 := \mathbb{E}_x \|g_{\mathcal{H}}(x)\|_2^2$.

Claim 4.3.2. *Consider any stage in the algorithm, with $\mathcal{H}$ being the set of functions at that stage, and $f \in F$ being the new function added to $\mathcal{H}$. Then*

- $0 \leq \|g_{\mathcal{H}}\|^2 \leq 1$;
- $\|g_{\mathcal{H}\cup\{f\}}\|^2 \geq \|g_{\mathcal{H}}\|^2 + \varepsilon^2$.

Proof. The first part of the claim is trivial as $g_{\mathcal{H}}$ maps to $P(Y)$. For the second part, observe that $\langle g_{\mathcal{H}\cup\{f\}} - g_{\mathcal{H}}, g_{\mathcal{H}}\rangle = 0$ and thus

$$\|g_{\mathcal{H}\cup\{f\}}\|_2^2 = \|g_{\mathcal{H}}\|_2^2 + \|g_{\mathcal{H}\cup\{f\}} - g_{\mathcal{H}}\|_2^2$$

We will show that $\|g_{\mathcal{H}\cup\{f\}} - g_{\mathcal{H}}\|_2^2 \geq \varepsilon^2$. We have

$$\begin{aligned}
\varepsilon &< |\mathbf{Pr}[g_{\mathcal{H}}(x) = f(x)] - \mathbf{Pr}[g(x) = f(x)]| \\
&= |\mathbb{E}_x\langle f(x), g_{\mathcal{H}}(x)\rangle - \mathbb{E}_x\langle f(x), g(x)\rangle| \\
&= \left|\mathbb{E}_x\langle f(x), g_{\mathcal{H}}(x)\rangle - \mathbb{E}_x\langle f(x), g_{\mathcal{H}\cup\{f\}}(x)\rangle\right| \qquad \text{(as } f \text{ is measurable with respect to } \mathcal{B}_{\mathcal{H}\cup\{f\}}\text{)} \\
&= \left|\mathbb{E}_x\langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H}\cup\{f\}}(x)\rangle\right| \\
&\leq \mathbb{E}_x\left|\langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H}\cup\{f\}}(x)\rangle\right|.
\end{aligned}$$

Now, as $f : X \to P(Y)$, for every $x \in X$, $\|f(x)\|_2 \leq 1$. Thus, by the Cauchy-Schwartz inequality, for every $x \in X$, we have

$$|\langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H}\cup\{f\}}(x)\rangle| \leq \|f(x)\|_2\|g_{\mathcal{H}\cup\{f\}}(x) - g_{\mathcal{H}}(x)\|_2 \leq \|g_{\mathcal{H}\cup\{f\}}(x) - g_{\mathcal{H}}(x)\|_2$$

Thus, by another application of the Cauchy-Schwartz inequality, we have

$$\varepsilon^2 \leq \mathbb{E}_x\left|\langle f(x), g_{\mathcal{H}}(x) - g_{\mathcal{H}\cup\{f\}}(x)\rangle\right|^2 \leq \|g_{\mathcal{H}\cup\{f\}} - g_{\mathcal{H}}\|_2^2.$$

□

This finishes the proof of the lemma. □

The following corollary for deterministic functions $f : X \to Y$ allows to obtain one-sided deterministic estimates. This simplifies some of the arguments later on.

Corollary 4.3.3. *Let $g : X \to Y$, $\varepsilon > 0$, and F be a collection of functions $f : X \to Y$. Then there exist $c \leq 1/\varepsilon^2$ functions $h_1, h_2, \ldots, h_c \in F$ such that for every $f \in F$, there is a function $\Gamma_f : Y^c \to Y$ such that*

$$\mathbf{Pr}_x[\Gamma_f(h_1(x), \ldots, h_c(x)) = f(x)] \geq \mathbf{Pr}_x[g(x) = f(x)] - \varepsilon.$$

Proof. Applying Lemma 4.3.1 to F we may assume the existence of $h_1, \ldots, h_c : X \to Y$ and $\Gamma : Y^C \to P(Y)$ such that for any $f \in F$,

$$|\mathbf{Pr}[f(x) = \Gamma(h_1(x), \ldots, h_c(x))] - \mathbf{Pr}[f(x) = g(x)]| \leq \varepsilon.$$

Let $A_{y_1,\ldots,y_c} = \{x \in X : h_1(x) = y_1, \ldots, h_c(x) = y_c\}$ be an atom defined by $h_1, \ldots, h_c$. Given $f \in F$, define $\Gamma_f : Y^c \to Y$ by letting $\Gamma_f(y_1, \ldots, y_c)$ to be the most common value that f attains on $A_{y_1,\ldots,y_c}$. Then

$$\begin{aligned}
&\mathbf{Pr}[f(x) = \Gamma_f(h_1(x), \ldots, h_c(x))] \\
&= \sum_{y_1,\ldots,y_c \in Y} \mathbf{Pr}[x \in A_{y_1,\ldots,y_c}] \cdot \max_{y^* \in Y} \mathbf{Pr}[f(x) = y^* | x \in A_{y_1,\ldots,y_c}] \\
&\geq \sum_{y_1,\ldots,y_c \in Y} \mathbf{Pr}[x \in A_{y_1,\ldots,y_c}] \cdot \mathbf{Pr}[f(x) = \Gamma(y_1, \ldots, y_c) | x \in A_{y_1,\ldots,y_c}] \\
&= \mathbf{Pr}[f(x) = \Gamma(h_1(x), \ldots, h_c(x))] \geq \mathbf{Pr}[f(x) = g(x)] - \varepsilon.
\end{aligned}$$

□

4.4 Proof of Theorem 5

Fix a prime field $\mathbb{F} = \mathbb{F}_p$. For $d \in \mathbb{N}$, we shorthand $\delta(d) = \delta_{\mathbb{F}}(d)$. We restate Theorem 5.

Theorem 5. *Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta(d) - \varepsilon) \leq c_{p,d,\varepsilon}.$$

We prove Theorem 5 in the remainder of this section. Let $g : \mathbb{F}^n \to \mathbb{U}_1$ be a received word where we identify $\mathbb{F}$ with $\mathbb{U}_1$. Apply Corollary 4.3.3 with $X =$

$\mathbb{F}^n$, $Y = \mathbb{U}_1$, $F = \mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{U}_1)$ and approximation parameter $\varepsilon/2$ to obtain $\mathcal{H} = \{h_1, \ldots, h_c\} \subseteq F$, $c \leq 4/\varepsilon^2$ such that, for every $f \in F$, there is a function $\Gamma_f : \mathbb{U}_1^c \to \mathbb{U}_1$ satisfying

$$\mathbf{Pr}[\Gamma_f(h_1(x), h_2(x), \ldots, h_c(x)) = f(x)] \geq \mathbf{Pr}[g(x) = f(x)] - \varepsilon/2.$$

Let $r_1, r_2 : \mathbb{N} \to \mathbb{N}$ be two non decreasing functions to be specified later, and let $C_{r,d}^{(6.4.6)}$ be as given in Lemma 6.4.6. We will require that for all $m \geq 1$,

$$r_1(m) \geq r_2(C_{r_2,d}^{(6.4.6)}(m+1)) + C_{r_2,d}^{(6.4.6)}(m+1) + 1. \tag{4.2}$$

As a first step, we r_1-regularize $\mathcal{H}$ by Lemma 6.4.6. This gives an r_1-regular factor $\mathcal{B}'$ of degree at most d, defined by polynomials $h'_1, \ldots, h'_{c'} : \mathbb{F}^n \to \mathbb{T}$, such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$, $c' \leq C_{r_1,d}^{(6.4.6)}(c)$ and $\mathrm{rank}(\mathcal{B}') \geq r_1(c')$. We denote $\mathcal{H}' = \{h'_1, \ldots, h'_{c'}\}$. Note that $\mathcal{H}'$ can have nonclassical polynomials as a result of the regularization. Let $\mathrm{depth}(h'_i) = k_i$ for $i \in [c']$. Let $G_f : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \to \mathbb{U}_1$ be defined such that

$$\Gamma_f(h_1(x), \ldots, h_c(x)) = G_f(h'_1(x), \ldots, h'_{c'}(x)).$$

Then

$$\mathbf{Pr}[G_f(h'_1(x), h'_2(x), \ldots, h'_{c'}(x)) = f(x)] \geq \mathbf{Pr}[g(x) = f(x)] - \varepsilon/2. \tag{4.3}$$

Next, given any classical polynomial $f : \mathbb{F}^n \to \mathbb{T}$ of degree at most d, we will show that if $\mathbf{Pr}[f(x) \neq g(x)] \leq \delta(d) - \varepsilon$, then f is measurable with respect to $\mathcal{H}'$ and this would upper bound the number of such polynomials by $p^{||\mathcal{B}'||} = p^{\prod_{i \in [c']} p^{k_i+1}}$ and as $c' = c'(p, d, \varepsilon)$ and $k_i \leq \left\lfloor \frac{d-1}{p-1} \right\rfloor$ this is independent on n.

Fix such a classical polynomial f. Appealing again to Lemma 6.4.6, we r_2-regularize $\mathcal{B}_f := \mathcal{B}' \cup \{f\}$. We get an r_2-regular factor $\mathcal{B}'' \succeq_{syn} \mathcal{B}'$ defined by the collection $\mathcal{H}'' = \{h'_1, \ldots, h'_{c'}, h''_1, \ldots, h''_{c''}\} \subseteq \mathrm{Poly}_{\leq d}(\mathbb{F}^n \to \mathbb{T})$. Note that it is a syntactic refinement of $\mathcal{B}'$ as by our choice of r_1,

$$\mathrm{rank}(\mathcal{B}') \geq r_1(c') \geq r_2(C^{(6.4.6)}_{r_2,d}(c'+1)) + C^{(6.4.6)}_{r_2,d}(c'+1) + 1 \geq r_2(|\mathcal{B}''|) + |\mathcal{B}''| + 1.$$

We will choose r_2 such that for all $m \geq 1$,

$$r_2(m) = \max\left(r_d^{(6.4.9)}\left(\frac{\varepsilon/4}{\left(p^{\lfloor \frac{d-1}{p-1} \rfloor + 1}\right)^m}\right), r_d^{(6.4.16)}(m)\right). \tag{4.4}$$

Let $\mathrm{depth}(h''_j) = l_j$ for $j \in [c'']$ and denote $S := \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \times \otimes_{j=1}^{c''} \mathbb{U}_{l_j+1}$. Since f is measurable with respect to $\mathcal{B}''$, there exists $F : S \to \mathbb{U}_1$ such that

$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)).$$

We next show that we can have each polynomial in the factor have a disjoint set of inputs, and still obtain more or less the same approximation factor.

Claim 4.4.1. *Let x^i, y^j, $i \in [c'], j \in [c'']$ be pairwise disjoint sets of $n \in \mathbb{N}$ variables each. Let $n' = n(c'+c'')$. Let $\tilde{f} : \mathbb{F}^{n'} \to \mathbb{U}_1$ and $\tilde{g} : \mathbb{F}^{n'} \to \mathbb{U}_1$ be defined as*

$$\tilde{f}(x) = F(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''}))$$

and

$$\tilde{g}(x) = G_f(h'_1(x^1), \ldots, h_{c'}(x^{c'})).$$

Then $\deg(\tilde{f}) \leq d$ *and*

$$\left|\mathbf{Pr}_{x \in \mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] - \mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))]\right| \leq \varepsilon/4.$$

Proof. The bound $\deg(\tilde{f}) \leq \deg(f) \leq d$ follows from Lemma 6.4.16 since $r_2(|\mathcal{H}''|) \geq r_d^{(6.4.16)}(|\mathcal{H}''|)$. To establish the bound on $\mathbf{Pr}[\tilde{f} = \tilde{g}]$, for each $s \in S$ let

$$p_1(s) = \mathbf{Pr}_{x \in \mathbb{F}^n}[(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)) = s].$$

Applying Lemma 6.4.9 and since our choice of r_2 satisfies $\mathrm{rank}(\mathcal{H}'') \geq r_d^{(6.4.9)}(\varepsilon/4|S|)$, we have that p_1 is nearly uniform over S,

$$p_1(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

Similarly, let

$$p_2(s) = \mathbf{Pr}_{x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''} \in \mathbb{F}^n}[(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''})) = s].$$

Note that the rank of the collection of polynomials $\{h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''})\}$ defined over $\mathbb{F}^{n'}$ cannot be lower than that of $\mathcal{H}''$. Applying Lemma 6.4.9 again gives

$$p_2(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

For $s \in S$, let $s' \in \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1}$ be the restriction of s to first c' coordinates, that is, $s' = (s_1, \ldots, s_{c'})$. Thus

$$\begin{aligned}
\mathbf{Pr}_{x \in \mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] &= \sum_{s \in S} p_2(s) 1_{F(s) = G_f(s')} \\
&= \sum_{s \in S} p_1(s) 1_{F(s) = G_f(s')} \pm \varepsilon/4 \\
&= \mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))] \pm \varepsilon/4.
\end{aligned}$$

□

So, we obtain that

$$\mathbf{Pr}_{x\in\mathbb{F}^{n'}}[\tilde{f}(x)=\tilde{g}(x)] \geq \mathbf{Pr}_{x\in\mathbb{F}^{n}}[f(x)=G_f(h'_1(x),\ldots,h'_{c'}(x))]-\varepsilon/4 \geq 1-\delta(d)+\varepsilon/4.$$

Next, we need the following variant of the Schwartz-Zippel lemma [146, 183].

Claim 4.4.2. *Let $d, n_1, n_2 \in \mathbb{N}$. Let $f_1 : \mathbb{F}^{n_1+n_2} \to \mathbb{F}$ and $f_2 : \mathbb{F}^{n_1} \to \mathbb{F}$ be such that* $\deg(f_1) \leq d$ *and*

$$\mathbf{Pr}[f_1(x_1,\ldots,x_{n_1+n_2}) = f_2(x_1,\ldots,x_{n_1})] > 1-\delta(d)$$

Then, f_1 does not depend on $x_{n_1+1},\ldots,x_{n_1+n_2}$.

Proof. We will show that f_1 does not depend on $z = x_{n_1+n_2}$ say. The proof for any other variable is similar. Recall that $\delta(d) := \frac{1}{p^a}\left(1-\frac{b}{p}\right)$ where $d = a\cdot(p-1)+b$. Let $f_1(x) = \sum_{k=0}^{d'} c_k z^k$ where $c_k \in \mathbb{F}[x_1,\ldots,x_{n_1+n_2-1}]$ and $d' \leq \min\{d, p-1\}$. Then $(f_1-f_2)(x) = c_0 - f_2(x) + \sum_{k=1}^{d'} c_k z^k$. We will show that $d' \geq 1$ will lead to a contradiction. Let $\deg(c_{d'}) = d''$. Note that $d''+d' \leq d$. Then,

$$\mathbf{Pr}[(f_1-f_2)(x)=0] \leq \mathbf{Pr}[c_{d'}=0] + (1-\mathbf{Pr}[c_{d'}=0])(1-\delta(d')) \leq 1-\delta(d'')\delta(d').$$

We will show that for any $d \geq 1$ and any $1 \leq c \leq p-1$, we have $\delta(c)\delta(d-c) \geq \delta(d)$ and this will show that $\mathbf{Pr}[(f_1-f_2)(x)=0] \leq 1-\delta(d'+d'') \leq 1-\delta(d)$ which leads to a contradiction. Thus, f_1 will not depend on z. We will now show that

$$\delta(c)\delta(d-c) \geq \delta(d) \tag{4.5}$$

Let $d = a\cdot(p-1)+b$.

Case 1: $0 \leq c \leq b$

$$\begin{aligned}(4.5) \quad &\Leftrightarrow \quad \left(1-\frac{c}{p}\right)\frac{1}{p^a}\left(1-\frac{b-c}{p}\right) \geq \frac{1}{p^a}\left(1-\frac{b}{p}\right) \\ &\Leftrightarrow \quad b \geq c\end{aligned}$$

Case 2: $b < c \leq p-1$

$$\begin{aligned}(4.5) \quad &\Leftrightarrow \quad \left(1-\frac{c}{p}\right)\frac{1}{p^{a-1}}\left(\frac{1+c-b}{p}\right) \geq \frac{1}{p^a}\left(1-\frac{b}{p}\right) \\ &\Leftrightarrow \quad (c-b)\left(1-\frac{c+1}{p}\right) \geq 0\end{aligned}$$

which is true by hypothesis. □

Now apply Claim 5.4.3 to $f_1 = \tilde{f}, f_2 = \tilde{g}, n_1 = nc', n_2 = nc''$. We obtain that $\tilde{f}$ does not depend on $y^1, \ldots, y^{c''}$. Hence,

$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = F(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), C_1, \ldots, C_{c''})$$

where $C_j = h''_j(0) \in \mathbb{U}_{l_j+1}$ for $j \in [c'']$. If we substitute $x^1 = \ldots = x^{c'} = x$ we get that

$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)) = F(h'_1(x), \ldots, h'_{c'}(x), C_1, \ldots, C_{c''}),$$

which shows that f is measurable with respect to $\mathcal{H}'$, as claimed.

4.5 Proof of Theorem 6

Theorem 6. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field. Let $\varepsilon > 0$ and $e \leq d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{F}}(d, n, \delta(e) - \varepsilon) \leq \exp\left(c_{p,d,\varepsilon} n^{d-e}\right).$$

The proof follows along the same lines as that of Theorem 5. It will rely on the following lemma which generalizes Claim 5.4.3.

Lemma 4.5.1. *Fix $d \geq e \geq 1, \varepsilon > 0$. There exists $r_{d,\varepsilon}^{(4.5.1)} \in \mathbb{N}$ such that the following holds. Let $f_1 : \mathbb{F}^{n_1+n_2} \to \mathbb{U}_1$ be a classical polynomial of degree at most d. Assume that*

- *There exist $f_2 : \mathbb{F}^{n_1} \to \mathbb{U}_1$ be such that $\mathbf{Pr}[f_1(x,y) = f_2(x)] \geq 1 - \delta(e) + \varepsilon$.*
- *There exists a polynomial $h : \mathbb{F}^{n_2} \to \mathbb{U}_{k+1}$ of degree at most d such that the factor it defines has rank at least $r_{d,\varepsilon}^{(4.5.1)}$, and a function $\Gamma : \mathbb{F}^{n_1} \times \mathbb{U}_{k+1} \to \mathbb{U}_1$, such that*

$$f_1(x,y) = \Gamma(x, h(y)).$$

- *The dependence on the depth of h is nontrivial: $f_1(x,y)$ cannot be written as $\Gamma'(x, p \cdot h(y))$ for any $\Gamma' : \mathbb{F}^{n_1} \times \mathbb{U}_k \to \mathbb{U}_1$.*

Then $\deg(h) \leq d - e$.

We first prove Theorem 6 assuming Lemma 4.5.1.

Proof of Theorem 6 assuming Lemma 4.5.1. The initial part of the proof is as in Theorem 5. Assume that $n > r_{d,\varepsilon/4}^{(4.5.1)}$ otherwise the theorem is trivially true. Let $f, g : \mathbb{F}^n \to \mathbb{U}_1$ with $\deg(f) \leq d$ and $\mathrm{dist}(f,g) \leq \delta(e) - \varepsilon$. For non decreasing functions $r_1, r_2 : \mathbb{N} \to \mathbb{N}$, chosen as in the proof of Theorem 5, we have an r_1-regular

$\mathcal{H}' = \{h'_1, \ldots, h'_{c'}\}$ and an r_2-regular $\mathcal{H}'' = \mathcal{H}' \cup \{h''_1, \ldots, h''_{c''}\}$ where each h'_i, h''_i is a nonclassical polynomial of degree $\leq d$, such that the following holds.

Let $\text{depth}(h'_i) = k_i$ for $i \in [c']$ and $\text{depth}(h''_j) = l_j$ for $j \in [c'']$. Since f is measurable with respect to $\mathcal{H}''$, there exists $F : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \times \otimes_{j=1}^{c''} \mathbb{U}_{l_j+1} \to \mathbb{U}_1$ such that

$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)).$$

We may assume that for all $i \in [c'']$, the depth of h''_i is minimal, in the sense that we cannot replace h''_i with $p \cdot h''_i$ and change F accordingly to still compute f (if this is not the case, then replace h''_i with $p \cdot h''_i$ whenever possible; this only reduces the degree of h''_i and the new factor has rank at least that of the original factor). Also, there exists a function $G_f : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \to \mathbb{U}_1$ such that

$$\mathbf{Pr}[G_f(h'_1(x), \ldots, h'_{c'}(x)) = f(x)] \geq 1 - \delta(e) + \varepsilon/2.$$

We will show that this implies that $\deg(h''_i) \leq d - e$ for all $i \in [c'']$. Let $\mathcal{B}''$ be the factor defined by $\mathcal{H}''$. As the number of polynomials of degree $d - e$ is exponential in n^{d-e}, the number of functions f is controlled by the product of the number of composing functions F, which is $p^{||\mathcal{B}''||} = p^{(\prod_{i \in [c']} p^{k_i+1})(\prod_{j \in [c'']} p^{l_j+1})} = c_1(p, d, \varepsilon)$, and the number of choices for $h''_1, \ldots, h''_{c''}$, which is $\exp(c_2 c'' n^{d-e})$. This amounts to at most $\exp(cn^{d-e})$ for some $c = c(p, d, \varepsilon)$, as claimed.

To prove the bound on the degrees of $h''_1, \ldots, h''_{c''}$, define, as in the proof of Theorem 5, x^i, y^j for $i \in [c'], j \in [c'']$ to be pairwise disjoint sets of $n \in \mathbb{N}$ variables. Let $n' = n(c' + c'')$. Define $\tilde{f} : \mathbb{F}^{n'} \to \mathbb{U}_1$ and $\tilde{g} : \mathbb{F}^{n'} \to \mathbb{U}_1$ as

$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = F(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''}(y^{c''}))$$

and

$$\tilde{g}(x^1, \ldots, x^{c'}) = G_f(h'_1(x^1), \ldots, h_{c'}(x^{c'})).$$

Then, by Claim 5.4.2, $\deg(\tilde{f}) \leq d$ and $\mathbf{Pr}[\tilde{f} = \tilde{g}] \geq 1 - \delta(e) + \varepsilon/4$.

We next apply Lemma 4.5.1 to show that $\deg(h''_j) \leq d - e$ for all $j \in [c'']$. To see that for say, $j = c''$, let $k = \mathrm{depth}(h''_{c''})$, $n_1 = n(c' + c'' - 1), n_2 = n, h(y) = h''_{c''}(y)$ and $\Gamma : \mathbb{F}^{n_1} \times \mathbb{U}_{k+1} \to \mathbb{U}_1$ given by

$$\Gamma((x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''-1}), \alpha) = F(h'_1(x^1), \ldots, h'_{c'}(x^{c'}), h''_1(y^1), \ldots, h''_{c''-1}(y^{c''-1}), \alpha).$$

so that

$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = \Gamma((x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''-1}), h''_{c''}(y^{c''})).$$

If we make sure that $r_2(m) \geq r^{(4.5.1)}_{d,\varepsilon/4}$ for all $m \geq 1$, then we establish all the requirements for Lemma 4.5.1. Hence we deduce that $\deg(h''_{c''}) \leq d - e$ as claimed.

□

4.5.1 Proof of Lemma 4.5.1

We prove Lemma 4.5.1 in this section. Fix $d \geq e \geq 1$ and $\varepsilon > 0$. Let $r = r^{(4.5.1)}_{d,\varepsilon}$ be large enough to be chosen later. We first show that we can replace h with a simple polynomial of the same degree and depth, which would allow us to simplify the analysis.

Let $\mathrm{depth}(h) = k$ and let $A = \deg(h) - (p-1)k$. Define $\tilde{h} : \mathbb{F}^{rA} \to \mathbb{U}_{k+1}$ as follows. Let $z = (z_{1,1}, \ldots, z_{r,A}) \in \mathbb{F}^{rA}$ and define

$$\tilde{h}(z) := \frac{\sum_{i=1}^{r} \prod_{j=1}^{A} z_{i,j}}{p^{k+1}}. \tag{4.6}$$

Note that $\tilde{h}$ and h are both polynomials of the same degree and depth. Define $\tilde{f}_1 : \mathbb{F}^{n_1+rA} \to \mathbb{U}_1$ as

$$\tilde{f}_1(x, z) = \Gamma(x, \tilde{h}(z)).$$

We will show that we may analyze $\tilde{f}_1$ instead of f_1 to obtain the upper bound on $\deg(h)$. To simplify the presentation, denote $Z_i := \prod_{j=1}^{A} z_{i,j}$ for $i \in [r]$. First, we argue that if r is chosen large enough then both $h, \tilde{h}$ are nearly uniform over $\mathbb{U}_{k+1}$.

Claim 4.5.2. *If r is chosen large enough then for all $\alpha \in \mathbb{U}_{k+1}$,*

$$\mathbf{Pr}_{y \in \mathbb{F}^{n_2}}[h(y) = \alpha] = p^{-(k+1)}(1 \pm \varepsilon/2)$$

and

$$\mathbf{Pr}_{z \in \mathbb{F}^{rA}}[\tilde{h}(z) = \alpha] = p^{-(k+1)}(1 \pm \varepsilon/2).$$

Proof. The proof for h follows from Lemma 6.4.9 by choosing $r \geq r_d^{6.4.9}\left(\frac{\varepsilon}{2p^{k+1}}\right)$. The proof for $\tilde{h}$ follows by a simple Fourier calculation. Let $\omega = \exp(2\pi i/p^{k+1})$. We have $\mathbf{Pr}[Z_i = 0], \mathbf{Pr}[Z_i = 1] \geq p^{-A} \geq p^{-d}$. One can verify that this implies that for any nonzero $c \in \mathbb{Z}_{p^{k+1}}$, $\mathbb{E}[\omega^{cZ_i}] \leq 1 - \eta$ for $\eta = p^{-O(d)}$. As $Z_1, \ldots, Z_r$ are independent we have $\mathbb{E}\left[\omega^{c(Z_1+\ldots+Z_r)}\right] \leq (1-\eta)^r$. Hence if we choose r large enough so that $(1-\eta)^r < (\varepsilon/2)p^{-(k+1)}$ then, for any $a \in \mathbb{Z}_{p^{k+1}}$,

$$\begin{aligned}\mathbf{Pr}[Z_1 + \ldots + Z_r = a \pmod{p^{k+1}}] &= p^{-(k+1)}\left(1 + \sum_{c \in \mathbb{Z}_{p^{k+1}} \setminus \{0\}} \omega^{-ac} \cdot \mathbb{E}\left[\omega^{c(Z_1+\ldots+Z_r)}\right]\right) \\ &= p^{-(k+1)}(1 \pm \varepsilon/2).\end{aligned}$$

□

This implies that $f_2(x)$ is also well approximates $\tilde{f}_1(x,z)$.

Corollary 4.5.3. $\mathbf{Pr}[\tilde{f}_1(x,z) = f_2(x)] \geq \mathbf{Pr}[f_1(x,y) = f_2(x)] - \varepsilon/2 \geq 1 - \delta(e) + \varepsilon/2$ *where* $x \in \mathbb{F}^{n_1}, y \in \mathbb{F}^{n_2}, z \in \mathbb{F}^{rA}$ *are chosen uniformly and independently.*

Proof. Claim 4.5.2 implies that the statistical distance between $h(y)$ and $\tilde{h}(z)$ is at most $\varepsilon/2$. Hence for every fixed x, $|\mathbf{Pr}[\Gamma(x,h(y)) = f_2(x)] - \mathbf{Pr}[\Gamma(x,\tilde{h}(z)) = f_2(x)]| \leq \varepsilon/2$. □

We next argue that by choosing r large enough, we can guarantee that $\tilde{f}_1$ has degree at most d.

Claim 4.5.4. *If* r *is chosen large enough then* $\deg(\tilde{f}_1) \leq \deg(f_1) \leq d$.

Proof. By Claim 4.5.2, if r is chosen large enough then $h(y), \tilde{h}(z)$ attain all possible values in $\mathbb{U}_{k+1}$. For every $\alpha \in \mathbb{U}_{k+1}$, let $f_\alpha(x) := \Gamma(x,\alpha)$. Note that as there exists some $y_\alpha \in h^{-1}(\alpha)$ then $f_\alpha(x) = f_1(x,y_\alpha)$ is a (classical) polynomial in x of degree at most d.

We have $f_1(x,y) = \Gamma(x,h(y)) = \Gamma'((f_\alpha(x) : \alpha \in \mathbb{U}_{k+1}), h(y))$ for some $\Gamma' : \mathbb{F}^{p^{k+1}} \times \mathbb{U}_{k+1} \to \mathbb{F}$. Let $\mathcal{H} = \{f_\alpha(x) : \alpha \in \mathbb{U}_{k+1}\}$ and for $r_1 : \mathbb{N} \to \mathbb{N}$ a growth function to be specified later, let $\mathcal{H}' = \{g_1(x), \ldots, g_c(x)\}$ be the result of r_1-regularizing $\mathcal{H}$ by Lemma 6.4.6. Then

$$f_1(x,y) = \Gamma''(g_1(x), \ldots, g_c(x), h(y))$$

for some $\Gamma'' : \mathbb{F}^c \times \mathbb{U}_{k+1} \to \mathbb{F}$. Hence also

$$\tilde{f}_1(x,z) = \Gamma(x,\tilde{h}(z)) = \Gamma''(g_1(x), \ldots, g_c(x), \tilde{h}(z)).$$

We next apply Lemma 6.4.16 to bound the degree of $\tilde{f}_1$. This requires to assume that $r_1(c) \geq r_d^{(6.4.16)}(c+1)$ and $r \geq r_d^{(6.4.16)}(C_{r_1,d}^{(6.4.6)}(p^{k+1})+1)$. We obtain that

$$\deg(\tilde{f}_1) = \deg(\Gamma''(g_1(x),\ldots,g_c(x),\tilde{h}(z))) \leq \deg(\Gamma''(g_1(x),\ldots,g_c(x),h(y))) = \deg(f_1) = d.$$

□

We next analyze the specific properties of $\tilde{h}$. Recall that we set $Z_i := \prod_{j=1}^{A} z_{i,j}$ so that $\tilde{h}(z) = \frac{\sum Z_i}{p^{k+1}}$. Since $\tilde{h}$ depends only on $W = \sum Z_i \mod p^{k+1}$, let the digits of $W \mod p^{k+1}$ in base p, be represented by classical polynomials $W_0(z),\ldots,W_k(z) : \mathbb{F}^{rA} \to \mathbb{F}$. Then, we can express $\tilde{f}_1(x,z)$ as

$$\tilde{f}_1(x,z) = \Gamma(x,\tilde{h}(z)) = \Gamma'(x,W_0(z),W_1(z),\ldots,W_k(z)) \tag{4.7}$$

for some $\Gamma' : \mathbb{F}^{n_1} \times \mathbb{F}^{k+1} \to \mathbb{U}_1$. Recall that we assumed that Γ depends nontrivially on the depth of its second argument. This implies that Γ' depends nontrivially on its last input (i.e. $W_k(z)$). As $\tilde{f}_1$ is a classical polynomial, and each W_i take values in $\mathbb{F}$, identifying $\mathbb{U}_1$ with $\mathbb{F}$, we can decompose

$$\tilde{f}_1(x,z) = \sum_{0 \leq d_0,\ldots,d_k \leq p-1} f_{d_0,\ldots,d_k}(x) \prod_{i=0}^{k} W_i(z)^{d_i}, \tag{4.8}$$

where $f_{d_0,\ldots,d_k} \in \mathbb{F}[x]$ is a classical polynomial. We next argue that $\deg(f_{d_0,\ldots,d_k})$ cannot be too large.

Lemma 4.5.5. $\deg(f_{d_0,\ldots,d_k}) \leq d - A\sum_{i=0}^{k} p^i d_i$ *for all* $0 \leq d_0,\ldots,d_k \leq p-1$.

We will require a few simple claims first. The ℓ-th symmetric polynomial in $Z = (Z_1,\ldots,Z_r)$, for $1 \leq \ell \leq r$, is a classical polynomial of degree ℓ defined as

$$S_\ell(Z) = \sum_{1 \leq i_1 < \ldots < i_\ell \leq r} \prod_{j=1}^{\ell} Z_{i_j}.$$

For $0 \leq i \leq k$, define $W'_i : \mathbb{F}^{rA} \to \mathbb{F}$ by $W'_i(z) := S_{p^i}(Z)$. The following claim follows immediately from Lucas theorem [123].

Claim 4.5.6. *Let $z \in \{0,1\}^{rA}$. Then, $W_i(z) = W'_i(z)$ for $i = 0, \ldots, k$.*

Proof. If $z \in \{0,1\}^{rA}$ then $Z \in \{0,1\}^r$. Lucas theorem implies that the i-th least significant digit (starting at 0) of $W = Z_1 + \ldots + Z_r$ in base p is given by $\binom{Z_1+\ldots+Z_r}{p^i} \mod p = S_{p^i}(Z)$. □

For every polynomial $P \in \mathbb{F}[z]$, define $\mathrm{ML}(P)$ to be the multilinearization of P. That is, it is obtained by replacing each $z_{i,j}^a$ by $z_{i,j}$ for all $a \geq 1$ and all $i \in [r], j \in [A]$. Note that $\mathrm{ML}(P)(z) = P(z)$ for all $z \in \{0,1\}^{rA}$.

Claim 4.5.7. *Let $P, Q : \mathbb{F}^{rA} \to \mathbb{F}$ be two polynomials such that $P(z) = Q(z)$ for all $z \in \{0,1\}^{rA}$. Then $\mathrm{ML}(P) \equiv \mathrm{ML}(Q)$.*

Proof. Let $n = rA$. It is easy to see that a multilinear polynomial $f : \mathbb{F}^n \to \mathbb{F}$ satisfies $f(z) = 0$ for all $z \in \{0,1\}^n$ if and only if $f \equiv 0$. Therefore, for every polynomial $P : \mathbb{F}^n \to \mathbb{F}$, $\mathrm{ML}(P)$ is the unique multilinear polynomial that agrees with P on $\{0,1\}^n$. Let $R : \mathbb{F}^n \to \mathbb{F}$ be defined as $R := P - Q$. Then by linearity, $\mathrm{ML}(R) :\equiv \mathrm{ML}(P) - \mathrm{ML}(Q)$. As $\mathrm{ML}(R) = 0$ for all $z \in \{0,1\}^n$, $\mathrm{ML}(R) \equiv 0$ which implies $\mathrm{ML}(P) \equiv \mathrm{ML}(Q)$. □

Proof of Lemma 4.5.5. For $D = \sum_{i=0}^{k} p^i d_i$, define

$$W^{(D)}(z) := \prod_{i=0}^{k} W_i(z)^{d_i}, \qquad W'^{(D)}(z) := \prod_{i=0}^{k} W'_i(z)^{d_i}.$$

By Claim 4.5.6 and Claim 4.5.7, we can define a common multilinearization of $W^{(D)}$ and $W'^{(D)}$ by

$$M^{(D)} := \mathrm{ML}\left(W^{(D)}\right) = \mathrm{ML}\left(W'^{(D)}\right).$$

Let $m'(z) = \prod_{i=1}^{D} Z_i = \prod_{i=1}^{D}\prod_{j=1}^{A} z_{i,j}$ be a monomial. The coefficient of m' in $W'^{(D)}$ is equal to the coefficient of $\prod_{i=1}^{D} Z_i$ in $\prod_{i=0}^{k} S_{p^i}(Z)^{d_i}$, which is equal to the number of partitions of a set of size D to d_0 sets of size 1, d_1 sets of size p, d_2 sets of size p^2, up to d_k sets of size p^k. This is given by

$$\prod_{i=0}^{k}\prod_{j=1}^{d_i}\binom{jp^i + d_{i+1}p^{i+1} + \ldots + d_k p^k}{p^i},$$

which by Lucas theorem is equal modulo p to $\prod_{i=0}^{k}(d_i!) \neq 0 \mod p$.

Owing to the above, we have $\deg(M^{(D)}) \leq \deg(W'^{(D)}) = AD$. Also, since $m'(z)$ is of maximal degree, it also remains in $M^{(D)}$ after multilinearization. Define

$$\bar{f}_1(x,z) := \sum_{0\leq d_0,\ldots,d_k\leq p-1} f_{d_0,\ldots,d_k}(x) M^{(D)}(z).$$

Then, we have $\deg(\bar{f}_1) \leq \deg(\tilde{f}_1) \leq d$.

Now, suppose that the lemma is false. Let $D = \sum p^i d_i$ be maximal such that $\deg(f_{d_0,\ldots,d_k}) > d - AD$. Note that D corresponds to a unique tuple $(d_0,\ldots,d_k)$. Let $m(x)$ be any monomial in $f_{d_0,\ldots,d_k}(x)$ with maximal degree, and recall that $m'(z) = \prod_{i=1}^{D} Z_i = \prod_{i=1}^{D}\prod_{j=1}^{A} z_{i,j}$. Hence, the monomial $m(x)m'(z)$, whose degree is larger than d, has a nonzero coefficient in $f_{d_0,\ldots,d_k}(x)M^{(D)}(z)$ as noted above. We will show it has a zero coefficient in any other $f_{d'_0,\ldots,d'_k}(x)M^{(D')}(z)$ with $(d'_0,\ldots,d'_k) \neq (d_0,\ldots,d_k)$, $D' = \sum_i p^i d'_i$ which will contradict the fact that $\deg(\bar{f}_1) \leq d$.

So, let $(d'_0, \ldots, d'_k) \neq (d_0, \ldots, d_k)$ and let $D' = \sum p^i d'_i$. Note that necessarily $D' \neq D$. If $D' > D$ then by maximality of D, $\deg\left(f_{d'_0,\ldots,d'_k}\right) \leq d - AD' < d - AD$ and hence $m(x)$ cannot appear in $f_{d'_0,\ldots,d'_k}(x)$. If $D' < D$ then $\deg\left(M^{(D')}\right) = AD' < AD$ and hence $m'(z)$ cannot appear in $M^{(D')}(z)$. □

Let $w = (w_0, \ldots, w_k) \in \mathbb{F}^{k+1}$ be new variables, and define $f'_1 : \mathbb{F}^{n_1+k+1} \to \mathbb{F}$ by

$$f'_1(x, w) = \Gamma'(x, w_0, \ldots, w_k) = \sum_{0 \leq d_0,\ldots,d_k \leq p-1} f_{d_0,\ldots,d_k}(x) \prod_{i=0}^{k} w_i^{d_i}. \tag{4.9}$$

We next argue that f'_1 is also well approximated by f_2.

Claim 4.5.8. $\mathbf{Pr}[f'_1(x, w) = f_2(x)] \geq \mathbf{Pr}[\tilde{f}_1(x, z) = f_2(x)] - \varepsilon/4 \geq 1 - \delta(e) + \varepsilon/4$, *where* $x \in \mathbb{F}^{n_1}, z \in \mathbb{F}^{rA}, w \in \mathbb{F}^{k+1}$ *are uniformly and independently distributed.*

Proof. By Claim 4.5.2, the distribution of $\tilde{h}$ is $\varepsilon/4$-close close in statistical distance to the uniform distribution over $\mathbb{U}_{k+1}$, hence the distribution of $(W_0(z), \ldots, W_k(z))$ is $\varepsilon/4$-close in statistical distance to the uniform distribution over $\mathbb{F}^{k+1}$. □

To conclude the proof of Lemma 4.5.1, expand $f'_1 - f_2$ as

$$f'_1(x, w) - f_2(x) = \sum_{i=0}^{d'} c_i(x, w_0, \ldots, w_{k-1}) w_k^i$$

where $c_i \in \mathbb{F}[x, w_0, \ldots, w_{k-1}]$, $d' \leq \min(d, p-1)$ and $c_{d'} \neq 0$. We have that $d' \geq 1$ since Γ' depends on $W_k(z)$. Also, by Lemma 4.5.5, for $i \geq 1$ we have $\deg(c_i) \leq d - Ap^k i$. To see this, suppose not. Consider the expansion in (4.9). Then, for some

$d_0, \ldots d_{k-1}$, $\deg(f_{d_0,\ldots d_{k-1},i}) + \sum_{j=0}^{k-1} d_j > d - Ap^k i$, which implies that

$$\deg(f_{d_0,\ldots d_{k-1},i}) > d - \sum_{j=0}^{k-1} d_j - Ap^k i \geq d - A\sum_{j=0}^{k-1} d_j p^j - Ap^k i,$$

which is a contradiction to Lemma 4.5.5. Hence

$$\begin{aligned}\mathbf{Pr}[f_1'(x,w) = f_2(x)] &\leq \mathbf{Pr}[c_{d'} = 0] + (1 - \mathbf{Pr}[c_{d'} = 0])(1 - \delta(d')) \\ &\leq 1 - \delta(d - Ap^k d')\delta(d') \leq 1 - \delta(d - d'(Ap^k - 1)),\end{aligned}$$

where the last inequality was established in Claim 5.4.3. So, as we established that $\delta(d - d'(Ap^k - 1)) < \delta(e)$ and $d' \geq 1$ we must have $Ap^k - 1 < d - e$, and hence $Ap^k \leq d - e$. Now, recall that $\deg(h) = \deg(\tilde{h}) = A + (p-1)k$ and it is a simple exercise to verify that $A + (p-1)k \leq Ap^k$ for all $A \geq 1, k \geq 0$. We thus showed that $\deg(h) \leq d - e$, as claimed.

4.6 Open Problems

Theorem 5 and Theorem 6 establish that over any fixed prime field $\mathbb{F}_p$ and any fixed $e \leq d$ and $\varepsilon > 0$, the number of degree d polynomials in a any ball of radius $\delta(e) - \varepsilon$ is at most $\exp(cn^{d-e})$ for some $c = c(p, d, \varepsilon)$, which in particular resolves the conjecture raised in [74] when $e = d$.

However, the bounds on c which we obtain are of Ackermann-type, which seem far from optimal. This leaves open the question of obtaining better bounds. In Chapter 6, we extend this to large prime fields and in Chapter 5 to fixed nonprime fields. However, the bounds are still weak in their dependence on the degree d. This may require a different approach, as currently higher-order Fourier analysis does not seem to provide better bounds.

Chapter 5

Higher order Fourier analysis over small nonprime fields with applications to list decoding and testing

5.1 Introduction

Fourier analysis over finite groups has played a central role in the development of theoretical computer science. Examples of its applications are everywhere: analysis of random walks on graphs [49], fast integer multiplication algorithms [145], learning algorithms [114], the Kahn-Kalai-Linial theorem [103], derandomization [128], tight inapproximability results using probabilistically checkable proofs [93], social choice theory [126], and coding theory [129]. See the surveys of De Wolf [50] and Štefankovič [152].

Higher-order Fourier analysis is a recent generalization of some aspects of Fourier analysis. Consider functions over the integers $\mathbb{Z}$. While classical Fourier analysis over $\mathbb{Z}$ studies correlations of functions with linear phases $e^{i\theta n}$, higher-order Fourier analysis over $\mathbb{Z}$ analyzes the correlation of functions with polynomial phases

such as $e^{i\theta n^2}$. The modern[1] work on higher-order Fourier analysis over $\mathbb{Z}$ began with the spectacular proof by Gowers of Szemerédi's theorem [77, 78], where the *Gowers norm* was introduced, and with the ergodic theory work of Host and Kra [96]. Subsequently, Green, Tao and Ziegler through several works [82, 83, 84, 85] largely completed the research program of understanding the relationships between different aspects of the theory over $\mathbb{Z}$. This work was applied to solve several longstanding open problems in additive number theory, including the celebrated result showing the existence of arbitarily long arithmetic progressions in the primes [83]. The book [161] by Tao on the subject surveys the current state of knowledge.

In an influential article [81], Green popularized the idea that it is useful to rephrase the problems arising in additive number theory into problems on vector spaces over fixed finite fields. The motivation was that many of the techniques in higher-order Fourier analysis over $\mathbb{Z}$ simplify over finite fields, because of the presence of subspaces and of algebraic notions such as orthogonality and linear independence. However, it was soon realized that these questions over finite fields are also intrinsically interesting because of their connections to theoretical computer science. In particular, the Gowers norm for functions on $\mathbb{F}^n$ for a finite prime field $\mathbb{F}$ is directly related to low-degree testing, a problem intensely studied by computer scientists since the early 90's.

Thanks to the sequence of works [80, 108, 163, 22, 164], the apparatus of higher-order Fourier analysis over $\mathbb{F}^n$ for any fixed prime order field $\mathbb{F}$ is also now

[1] In retrospect, Weyl's results on equidistribution of polynomial phases [173] laid the foundations of this theory.

largely complete. The theory has subsequently found several interesting applications in computer science that we detail below and has become part of the mainstream theorist toolkit. However, in all of these applications, the finite field in consideration was restricted to be a field of *prime* order, while the problems themselves are interesting over general finite fields. In this work, we show how the techniques of higher-order Fourier analysis continue to apply even when the underlying field is a non-trivial extension of a prime order field.

5.1.1 Applications

In this section, we describe three different problems involving a finite field $\mathbb{K}$, which previously had been solved only when $|\mathbb{K}|$ was prime but which we can now solve for arbitrary finite $\mathbb{K}$.

Throughout, let $\mathbb{F}$ be a fixed prime order field, and let $\mathbb{K}$ be a finite field that extends $\mathbb{F}$. Let $q = |\mathbb{K}|$, $p = |\mathbb{F}|$ and $q = p^r$ for $r > 0$.

List-decoding Reed-Muller codes

The notion of *list decoding* was introduced by Elias [60] and Wozencraft [176] to decode *error correcting codes* beyond half the minimum distance. The goal of a list decoding algorithm is to produce all the codewords within a specified distance from the received word. At the same time one has to find the right radius for which the number of such codewords is small, otherwise there is no hope for the algorithm to be efficient. After the seminal results of Goldreich and Levin [67] and Sudan [153] which gave list decoding algorithms for the Hadamard code and the Reed-Solomon code respectively, there has been tremendous progress in designing list decodable

codes. See the survey by Guruswami [89, 88] and Sudan [154].

List decoding has applications in many areas of computer science including hardness amplification in complexity theory [155, 165], derandomization [168], construction of hard core predicates from one way functions [67, 2], construction of extractors and pseudorandom generators [157, 149] and computational learning [116, 101]. However, the largest radius up to which list decoding is tractable is still a fundamental open problem even for well studied codes like Reed-Solomon (univariate polynomials) and Reed-Muller codes (multivariate polynomials). The goal of this work is to analyse Reed-Muller codes over small fields (possible non prime) and small degree.

Reed-Muller codes (RM codes) were introduced in Chapter 4. We recap the essential details here again. Reed-Muller codes (RM codes) were discovered by Muller in 1954. Let $d \in \mathbb{N}$. The RM code $\mathrm{RM}_{\mathbb{K}}(n, d)$ is defined as follows. The message space consists of degree $\leq d$ polynomials in n variables over $\mathbb{K}$ and the codewords are evaluation of these polynomials on $\mathbb{K}^n$. Let $\delta_q(d)$ denote the normalized distance of $\mathrm{RM}_{\mathbb{K}}(n, d)$. Let $d = a(q-1) + b$ where $0 \leq b < q - 1$. We have

$$\delta_{\mathbb{K}}(d) = \frac{1}{q^a}\left(1 - \frac{b}{q}\right).$$

RM codes are one of the most well studied error correcting codes. Many applications in computer science involve low degree polynomials over small fields, namely RM codes. Given a received word $g : \mathbb{K}^n \to \mathbb{K}$ the objective is to output the list of codewords (e.g. low-degree polynomials) that lie within some distance of

g. Typically we will be interested in regimes where list size is either independent of n or polynomial in the block length q^n.

Let $\mathcal{P}_d(\mathbb{K}^n)$ denote the class of degree $\leq d$ polynomials $f : \mathbb{F}^n \to \mathbb{F}$. Let dist denote the normalized Hamming distance. For $\mathrm{RM}_{\mathbb{K}}(n, d)$, $\eta > 0$, let

$$\ell_{\mathbb{F}}(n, d, \eta) := \max_{g:\mathbb{F}^n \to \mathbb{F}} |\{f \in \mathcal{P}_d(\mathbb{F}^n) : \mathrm{dist}(f, g) \leq \eta\}| .$$

Let $\mathrm{LDR}_{\mathbb{K}}(n, d)$ (short for *list decoding radius*) be the maximum ρ for which $\ell_{\mathbb{K}}(n, d, \rho - \varepsilon)$ is upper bounded by a constant depending only on $\varepsilon, |\mathbb{K}|, d$ for all $\varepsilon > 0$.

It is easy to see that $\mathrm{LDR}_{\mathbb{K}}(n, d) \leq \delta_{\mathbb{K}}(d)$. The difficulty lies in proving a matching lower bound. We review some previous work next. The first breakthrough result was the celebrated work of Goldreich and Levin [67] who showed that in the setting of $d = 1$ over $\mathbb{F}_2$ (Hadamard Codes) $\mathrm{LDR}_{\mathbb{F}_2}(n, 1) = \delta_{\mathbb{F}_2}(1) = 1/2$. Later, Goldreich, Rubinfield and Sudan [68] generalized the field to obtain $\mathrm{LDR}_{\mathbb{K}}(n, 1) = \delta_{\mathbb{K}}(1) = 1 - 1/|\mathbb{K}|$. In the setting of $d < |\mathbb{K}|$, Sudan, Trevisan and Vadhan [155] showed that $\mathrm{LDR}_{\mathbb{K}}(n, d) \geq 1 - \sqrt{2d/|\mathbb{K}|}$ improving previous work by Arora and Sudan [5], Goldreich *et al* [68] and Pellikaan and Wu [133]. Note that this falls short of the upper bound which is $\delta_{\mathbb{K}}(d)$.

In 2008, Gopalan, Klivans and Zuckerman [74] showed that $\mathrm{LDR}_{\mathbb{F}_2}(n, d) = \delta_{\mathbb{F}_2}(d)$.They posed the following conjecture.

Conjecture 3 ([74])**.** *For fixed d and finite field $\mathbb{K}$, $\mathrm{LDR}_{\mathbb{K}}(n, d) = \delta_{\mathbb{K}}(d)$.*

It is believed [74, 73] that the hardest case is the setting of small d. An important step in this direction was taken in [73] that considered quadratic poly-

nomials and showed that $\mathrm{LDR}_{\mathbb{K}}(n,2) = \delta_{\mathbb{K}}(2)$ for all fields $\mathbb{K}$ and thus proved the conjecture for $d = 2$. In Chapter 4, we resolved the conjecture for prime $\mathbb{K}$.

Our main result for list decoding is a resolution of Conjecture 3.

Theorem 5.1.1. *Let $\mathbb{K}$ be a finite field. Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,*

$$\ell_{\mathbb{K}}(d, n, \delta_{\mathbb{K}}(d) - \varepsilon) \leq c_{|\mathbb{K}|,d,\varepsilon}.$$

Thus,

$$\mathrm{LDR}_{\mathbb{K}}(n,d) = \delta_{\mathbb{K}}(d).$$

Remark 5.1.2 (Algorithmic Implications)**.** *Using the blackbox reduction of algorithmic list decoding to combinatorial list decoding in [74] along with Theorem 5.1.1, for fixed finite fields, d and $\varepsilon > 0$, we now have list decoding algorithms in both the global setting (running time polynomial in $|\mathbb{K}|^n$) and the local setting (running time polynomial in n^d).*

Algorithmic polynomial decomposition

Consider the following family of properties of functions over a finite field $\mathbb{K}$.

Definition 5.1.3. *Given a positive integer k, a vector of positive integers $\mathbf{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and a function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, we say that a function $P : \mathbb{K}^n \to \mathbb{K}$ is* $(k, \mathbf{\Delta}, \Gamma)$-structured *if there exist polynomials $P_1, P_2, \ldots, P_k : \mathbb{K}^n \to \mathbb{K}$ with each $\deg(P_i) \leq \Delta_i$ such that for all $x \in \mathbb{K}^n$,*

$$P(x) = \Gamma(P_1(x), P_2(x), \ldots, P_k(x)).$$

The polynomials $P_1, \ldots, P_k$ are said to form a $(k, \mathbf{\Delta}, \Gamma)$-decomposition.

For instance, an n-variate polynomial over the field $\mathbb{K}$ of total degree d factors nontrivially exactly when it is $(2, (d-1, d-1), \mathsf{prod})$-structured where $\mathsf{prod}(a, b) = a \cdot b$. We shall use the term *degree-structural property* to refer to a property from the family of $(k, \mathbf{\Delta}, \Gamma)$-structured properties.

The problem here is, for arbitrary fixed $k, \mathbb{K}, (\Delta), \Gamma$, given a polynomial, decide efficiently if it is degree structural and if yes, output the decomposition. An efficient algorithm for the above would imply a (deterministic) poly(n)-time algorithm for factoring an n-variate polynomial of degree d over $\mathbb{K}$. Also, it implies a polynomial time algorithm for deciding whether a d-dimensional tensor over $\mathbb{K}$ has rank at most r.

Also, it would give polynomial time algorithms for a wide range of problems not known to have non-trivial solutions previously, such as whether a polynomial of degree d can be expressed as $P_1 \cdot P_2 + P_3 \cdot P_4$ where each P_1, P_2, P_3, P_4 are of degree $d-1$ or less.

This problem was solved for prime $\mathbb{K}$, satisfying $d < |\mathbb{F}|$ by Bhattacharyya [23] and later for all d and prime $|\mathbb{K}|$ by Bhattacharyya, Hatami and Tulsiani [30].

Our main result in this line of work establishes this for all finite fields.

Theorem 5.1.4. *For every positive integer k, every vector of positive integers $\mathbf{\Delta} = (\Delta_1, \Delta_2, \ldots, \Delta_k)$ and every function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, there is a deterministic algorithm $\mathcal{A}_{k,\mathbf{\Delta},\Gamma}$ that takes as input a polynomial $P : \mathbb{K}^n \to \mathbb{K}$ of degree d that runs in time polynomial in n, and outputs a $(k, \mathbf{\Delta}, \Gamma)$-decomposition of P if one exists while otherwise returning* NO.

Testing affine-invariant properties

The goal of property testing, as initiated by [38, 8] and defined formally by [139, 69], is to devise algorithms that query their input a very small number of times while correctly deciding whether the input satisfies a given property or is "far" from satisfying it. A property is called *testable* if the query complexity can be made independent of the size of the input.

More precisely, we use the following definitions. Let $[R]$ denote the set $\{1, \ldots, R\}$. Given a property $\mathcal{P}$ of functions in $\{\mathbb{K}^n \to [R] \mid n \in \mathbb{Z}_{\geq 0}\}$, we say that $f : \mathbb{K}^n \to [R]$ is *ε-far* from $\mathcal{P}$ if

$$\min_{g \in \mathcal{P}} \mathbf{Pr}_{x \in \mathbb{K}^n}[f(x) \neq g(x)] > \varepsilon,$$

and we say that it is *ε-close* otherwise.

Definition 5.1.5 (Testability)**.** *A property $\mathcal{P}$ is said to be* testable *(with one-sided error) if there are functions $q : (0,1) \to \mathbb{Z}_{>0}$, $\delta : (0,1) \to (0,1)$, and an algorithm T that, given as input a parameter $\varepsilon > 0$ and oracle access to a function $f : \mathbb{K}^n \to [R]$, makes at most $q(\varepsilon)$ queries to the oracle for f, always accepts if $f \in \mathcal{P}$ and rejects with probability at least $\delta(\varepsilon)$ if f is ε-far from $\mathcal{P}$. If, furthermore, q is a constant function, then $\mathcal{P}$ is said to be* proximity-obliviously testable (PO testable)*.*

The term proximity-oblivious testing is coined by Goldreich and Ron in [72]. As an example of a testable (in fact, PO testable) property, let us recall the famous result by Blum, Luby and Rubinfeld [38] which initiated this line of research. They showed that linearity of a function $f : \mathbb{K}^n \to \mathbb{K}$ is testable by a test which makes 3

queries. This test accepts if f is linear and rejects with probability $\Omega(\varepsilon)$ if f is ε-far from linear.

Linearity, in addition to being testable, is also an example of a *linear-invariant* property. We say that a property $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ is linear-invariant if it is the case that for any $f \in \mathcal{P}$ and for any $\mathbb{K}$-linear transformation $L : \mathbb{K}^n \to \mathbb{K}^n$, it holds that $f \circ L \in \mathcal{P}$. Similarly, an *affine-invariant* property is closed under composition with affine transformations $A : \mathbb{K}^n \to \mathbb{K}^n$ (an affine transformation A is of the form $L + c$ where L is $\mathbb{K}$-linear and $c \in \mathbb{K}$). The property of a function $f : \mathbb{K}^n \to \mathbb{K}$ being affine is testable by a simple reduction to [38], and is itself affine-invariant. Other well-studied examples of affine-invariant (and hence, linear-invariant) properties include Reed-Muller codes [8, 7, 61, 139, 4] and Fourier sparsity [76]. In fact, affine invariance seems to be a common feature of most interesting properties that one would classify as "algebraic". Kaufman and Sudan in [110] made explicit note of this phenomenon and initiated a general study of the testability of affine-invariant properties (see also [71]). Our main result for testing is the following:

Theorem 5.1.6 (Main testing result)**.** *Let $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ be an affine-invariant property, where R and* $\mathrm{char}(\mathbb{K})$ *are fixed positive integers. Then, $\mathcal{P}$ is PO testable with t queries if and only if $\mathcal{P}$ is t, W-lightly locally characterized for some $W > 0$.*

The above result has an extra restriction than usual testing results in the form of a weight restriction. This is required when the field size is growing. We explain this in the next section. In the fixed field case, we have a testing theorem without any extra conditions, simply because W is trivially bounded.

Theorem 5.1.7 (Main testing result-Fixed fields)**.** *Let $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ be an affine-invariant property, where R and* $\mathrm{char}(\mathbb{K})$ *are fixed positive integers. Then, $\mathcal{P}$ is PO testable with t queries if and only if $\mathcal{P}$ is t-locally characterized.*

A property $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R] : n \geq 1\}$ is said to be t-locally characterized if there are t affine forms $L_1, \ldots, L_t$ on $\ell \leq t$ variables such that $f \in \mathcal{P}$ if and only if there exist $x_1, \ldots, x_\ell \in \mathbb{K}^n$ with $(f(L_1(x_1, \ldots, x_\ell)), \ldots, f(L_t(x_1, \ldots, x_\ell)))$ satisfying a particular constraint. Therefore, a natural test for a t-locally characterized property is to choose random $x_1, \ldots, x_\ell$, evaluate f on $L_1(x_1, \ldots, x_\ell), \ldots, L_t(x_1, \ldots, x_\ell)$ and check whether the evaluations satisfy the constraint. Indeed, this is the test we analyze.

A very interesting feature of Theorem 5.1.7 is that $\mathbb{K}$ is allowed to be growing, which is not true in our other two applications. Setting $R = 2$, Theorem 5.1.7 shows that any affine-invariant property $\mathcal{P}$ of subsets of $\mathbb{K}^n$ are PO testable if the property is locally characterized (with respect to $\mathbb{K}$-affine constraints). Previously, [27] proved Theorem 5.1.7 in the case that $\mathbb{K}$ is of fixed prime order using higher-order Fourier analytic techniques. We note that other recent results on 2-sided testability of affine-invariant properties over fixed prime-order fields [94, 181] can also be similarly extended to non-prime fields but we omit their description here.

5.1.2 Our Techniques

New Ingredients

Our starting point is the observation that $\mathbb{K}$ is an r-dimensional vector space over $\mathbb{F}$. Thus, we can view a function $Q : \mathbb{K}^n \to \mathbb{K}$ as determined by a collection

of functions $P_1, \ldots, P_r : \mathbb{K}^n \to \mathbb{F}$ where $\mathbb{K}^n$ is viewed as $\mathbb{F}^{rn}$. In view of this, we define the notion of an *additive polynomial*. A function[2] $P : \mathbb{K}^n \to \mathbb{F}$ is said to have *additive degree* d if for all $h_1, \ldots, h_{d+1} \in \mathbb{K}^n$, $D_{h_1} \cdots D_{h_{d+1}} P \equiv 0$, where $(D_h P)(x) = P(x+h) - P(x)$. Additive polynomials are exactly the non-classical polynomials of [164] when the domain is $\mathbb{F}^{rn}$. Moreover, if $Q : \mathbb{K}^n \to \mathbb{K}$ has degree d (in the usual sense of having a monomial with degree d), then $\mathrm{Tr}(\alpha Q)$ has additive degree $\leq d$ for any $\alpha \in \mathbb{K}$ where $\mathrm{Tr} : \mathbb{K} \to \mathbb{F}$ denotes the trace function.

Therefore, we can directly write any polynomial $P : \mathbb{K}^n \to \mathbb{K}$ in terms of additive polynomials and then import all of the results shown in [164] for non-classical polynomials to our setting! Unfortunately, we are not done. The reason is that our applications require, in addition to additive structure, some of the multiplicative structure of $\mathbb{K}$, which is lost when we view $\mathbb{K}$ as $\mathbb{F}^r$.

To see why, recall the question of testing affine-invariant properties. As mentioned above, we can view any one-sided test as examining the restriction of the input function on a random K-dimensional affine subspace of $\mathbb{K}^n$, for some constant integer K. In other words, the test will evaluate the input function at elements of the set $H = \{x + \sum_{i=1}^{K} a_i y_i : a_1, \ldots, a_K \in \mathbb{K}\}$ for some $x, y_1, \ldots, y_K \in \mathbb{K}$. Clearly, H is not an affine subspace of $\mathbb{F}^{rn}$. An important component of the higher-order Fourier analytic approach is to show that any "sufficiently pseudorandom" collection of polynomials is equidistributed on H, and the proof of this fact in [27] crucially uses that H is a subspace of a vector space over a prime field. In our work, we show

[2]To deal with low characteristics, we will actually use a slightly general definition valid for functions mapping to the torus $\mathbb{R}/\mathbb{Z}$.

a strong equidistribution theorem (Theorem 5.3.3) that holds when H is an affine subspace of $\mathbb{K}^n$.

A different place where multiplicative structure rears its head is a key *Degree Preserving Lemma* of [27]. Informally, it states that if $P_1, \ldots, P_C$ form a "sufficiently pseudorandom" collection of polynomials and $F(x) = \Gamma(P_1(x), \ldots, P_C(x))$ is a polynomial of degree d where Γ is an arbitrary composition function, then for any other collection of polynomials $Q_1, \ldots, Q_C$ where $\deg(Q_i) \leq \deg(P_i)$ for every i, $G(x) = \Gamma(Q_1(x), \ldots, Q_C(x))$ also has degree $\leq d$. The lemma is crucially used for the analysis of the Reed-Muller list decoding bound in [35] and the polynomial decomposition algorithm in [23, 30]. Its proof goes via showing that if all $(d+1)$ iterated derivatives of $F : \mathbb{K}^n \to \mathbb{K}$ vanish, then so must all $(d+1)$ iterated derivatives of $G : \mathbb{K}^n \to \mathbb{K}$. However, when $|\mathbb{K}|$ is non-prime, all $(d+1)$ iterated derivates of a function $G : \mathbb{K}^n \to \mathbb{K}$ may vanish without the degree being $\leq d$; consider for example the polynomial x^p which vanishes after only 2 derivatives.

We resolve this issue by giving a different and more transparent proof of the Degree Preserving Lemma, which actually holds in a much more general setting (Theorem 5.3.4). Using the above notation, we prove that if $F : \mathbb{K}^n \to \mathbb{K}$ satisfies some locally characterized property $\mathcal{P}$, then $G : \mathbb{K}^n \to \mathbb{K}$ does also. Since due to [109], we know that degree is locally characterized, our desired result follows. Our new proof uses our strong equidistribution theorem on affine subspaces of $\mathbb{K}^n$.

An interesting point to note is that both the equidistribution theorem and the degree preserving lemma work only assuming that the field characteristic is constant, without any assumption on the field size.

Reed-Muller codes

For a received word $g : \mathbb{K}^n \to \mathbb{K}$ our goal is to upper bound $|\{f \in \mathcal{P}_d : \text{dist}(f, g) \leq \eta\}|$, where $\eta = \delta_{\mathbb{K}}(d) - \varepsilon$ for some $\eta > 0$ and $\mathcal{P}_d$ is the class $\{Q : \mathbb{K}^n \to \mathbb{K} : \deg(Q) \leq d\}$. The proof technique is similar in structure as [35]. We apply the weak regularity lemma (Corollary 5.4.1) to the received word $g : \mathbb{K}^n \to \mathbb{K}$ and reduce the problem to a structured word $g' : \mathbb{K}^n \to \mathbb{K}$. More specifically, whenever $\text{dist}(f, g) \leq \eta$, we have $\text{dist}(f, g') \leq \eta + \varepsilon/2$. From here, we first express each function $f : \mathbb{K}^n \to \mathbb{K}$ as a linear combination of functions $f' : \mathbb{K}^n \to \mathbb{F}$. It can be then shown that the analysis in [35] works for functions $f' : \mathbb{K}^n \to \mathbb{F}$. A naive recombination of the $f' : \mathbb{K}^n \to \mathbb{F}$ to $f : \mathbb{K}^n \to \mathbb{K}$ gives us useful bounds only when $d < \text{char}(|\mathbb{F}|)$. To circumvent this problem, we use our improved degree preserving theorem. This is crucial to our analysis as the technique of [35] can be used only to analyze the additive degree of polynomials which is not enough for the argument to work for arbitrary d and $|\mathbb{K}|$.

Polynomial decomposition

The algorithm and its analysis follows the lines of [23, 30]. Given a polynomial $P : \mathbb{K}^n \to \mathbb{K}$ (where $|\mathbb{K}|$ is bounded), we consider the collection of additive polynomials $\{\text{Tr}(\alpha_1 P), \ldots, \text{Tr}(\alpha_r P)\}$ where $\alpha_1, \ldots, \alpha_r \in \mathbb{K}$ are linearly independent. We regularize this collection into a pseudorandom additive polynomial factor and set one variable to 0 such that the degrees of the polynomials do not change. We then recursively solve the problem on $n - 1$ variables and then apply a lifting procedure to get a decomposition for the original problem. A naive analysis of the lifting procedure over non-prime fields requires that $\deg(P) < \text{char}(\mathbb{F})$. In order to

get around this, we use our improved degree preserving theorem which applies for arbitrary degrees.

Testing affine-invariant properties

Suppose $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ is a locally characterized affine-invariant property (where R and char($\mathbb{K}$) are bounded but $n|\mathbb{K}|$ is growing). Our proof follows the lines of [29, 28, 27]. Suppose f is far from $\mathcal{P}$. We first identify a low-rank function close to f in an appropriate Gowers norm which also contains the violation that f contains. Here, low rank is with respect to a collection $\mathcal{B}$ of *additive* polynomials. We then investigate the distribution of $\mathcal{B}$ on the affine constraint that f violates. Since these are affine with respect to $\mathbb{K}^n$, we need to use our strong equidistribution theorem. The rest of the proof proceeds along the same lines as [27].

Because the proof of Theorem 5.1.7 is very analogous to that in [27] (except for the use of additive polynomials and the new equidistribution theorem) and requires significant additional notation, we omit it here.

5.2 Preliminaries

Let $\mathbb{N}$ denote the set of positive integers. For $n \in \mathbb{N}$, let $[n] := \{1, 2, \ldots, n\}$. We use $y = x \pm \varepsilon$ to denote $y \in [x - \varepsilon, x + \varepsilon]$. For $n \in \mathbb{N}$, and $x, y \in \mathbb{C}^n$, let $\langle x, y\rangle := \sum_{i=1}^n x_i \overline{y_i}$ where $\overline{a}$ is the conjugate of a. Let $\|x\|_2 := \sqrt{\langle x, x\rangle}$.

Let $\mathbb{T}$ denote the torus $\mathbb{R}/\mathbb{Z}$. This is an abelian group under addition. Let $e : \mathbb{T} \to \mathbb{C}$ be the function $e(x) = e^{2\pi i x}$. For an integer $k \geq 0$, let $\mathbb{U}_k := \frac{1}{p^k}\mathbb{Z}/\mathbb{Z}$. Note that $\mathbb{U}_k$ is a subgroup of $\mathbb{T}$. Let $\iota : \mathbb{F} \to \mathbb{U}_1$ be the bijection $\iota(a) = \frac{|a|}{p} \pmod 1$.

Fix a prime field $\mathbb{F} = \mathbb{F}_p$, and let $\mathbb{K} = \mathbb{F}_q$ where $q = p^r$ for a positive integer r. Also, fix basis $\alpha_1, \ldots, \alpha_r \in \mathbb{K}$ such that $\mathbb{K} = \{c_1\alpha_1 + c_2\alpha_2 + \cdots + c_r\alpha_r : c_1, \ldots, c_r \in \mathbb{F}\}$. We denote by $\mathrm{Tr} : \mathbb{K} \to \mathbb{F}$ the trace function:

$$\mathrm{Tr}(x) = x + x^p + x^{p^2} + \cdots + x^{p^{r-1}}$$

Recall that $\{x \to \mathrm{Tr}(ax) : a \in \mathbb{K}\}$ is in bijection with the set of all linear maps from $\mathbb{K}$ to $\mathbb{F}$. Also, we use $|\cdot|$ to denote the obvious map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$.

We will need the following useful fact.

Proposition 5.2.1. *There exist $\alpha_1, \alpha_2, \ldots, \alpha_r$ in $\mathbb{K}$ such that any $x \in \mathbb{K}$ equals $\sum_{i=1}^{r} \alpha_i \mathrm{Tr}(\alpha_i x)$.*

5.2.1 Affine forms and constraints

A *linear form on k variables* is a vector $L = (w_1, w_2, \ldots, w_k) \in \mathbb{K}^k$ that is interpreted as a function from $(\mathbb{K}^n)^k$ to $\mathbb{K}^n$ via the map $(x_1, \ldots, x_k) \mapsto w_1x_1 + w_2x_2 + \cdots + w_kx_k$. A linear form $L = (w_1, w_2, \ldots, w_k)$ is said to be *affine* if $w_1 = 1$. From now, linear forms will always be assumed to be affine.

We specify a partial order $\preceq$ among affine forms. We say $(w_1, \ldots, w_k) \preceq (w'_1, \ldots, w'_k)$ if $|\mathrm{Tr}(\alpha_j w_i)| \leq |\mathrm{Tr}(\alpha_j w'_i)|$ for all $i \in [k], j \in [r]$. An affine constraint is a collection of affine forms, with the added technical restriction of being downward-closed with respect to $\preceq$. For future references we state this as the following definition.

Definition 5.2.2 (Affine constraints)**.** *An* affine constraint of size m on k variables *is a tuple $A = (L_1, \ldots, L_m)$ of m affine forms $L_1, \ldots, L_m$ over $\mathbb{F}$ on k variables, where:*

(i) $L_1(x_1, \ldots, x_k) = x_1$;

(ii) If L *belongs to* A *and* $L' \preceq L$, *then* L' *also belongs to* A.

5.2.2 Polynomials, Degrees and Derivatives

A function $P : \mathbb{K}^n \to \mathbb{K}$ is a *polynomial of degree* d if for all $d_1, \ldots, d_n \geq 0$ such that $\sum_i d_i \leq d$, there exists $c_{d_1,\ldots,d_n} \in \mathbb{K}$ such that:

$$P(x_1, \ldots, x_n) = \sum_{\substack{d_1,\ldots,d_n \in \mathbb{Z}^+: \\ d_1+\cdots+d_n \leq d}} c_{d_1,\ldots,d_n} x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n}$$

We use the notion of *additive degree* for functions mapping to $\mathbb{T}$. Given a function $f : \mathbb{K}^n \to \mathbb{T}$, its *additive derivative in direction* $h \in \mathbb{K}^n$ is $D_h f : \mathbb{K}^n \to \mathbb{T}$, given by

$$D_h f(x) = f(x + h) - f(x).$$

Definition 5.2.3 (Additive Polynomials)**.** *A function* $P : \mathbb{K}^n \to \mathbb{T}$ *is a* polynomial of additive degree d *if for all* $x, h_1, h_2, \ldots, h_{d+1} \in \mathbb{K}^n$, *we have*

$$D_{h_1} D_{h_2} \cdots D_{h_{d+1}} P(x) = 0. \tag{5.1}$$

A function of bounded additive degree is called an additive polynomial.

For functions P mapping to $\mathbb{T}$, $\deg(P)$ denotes its additive degree. Note that we can interpret $P : \mathbb{K}^n \to \mathbb{T}$ as a function $P' : \mathbb{F}^{nr} \to \mathbb{T}$ with the same additive degree by setting $P(x_1, \ldots, x_n) = P'(\mathrm{Tr}(\alpha_1 x_1), \ldots, \mathrm{Tr}(\alpha_r x_1), \ldots, \mathrm{Tr}(\alpha_1 x_n), \ldots, \mathrm{Tr}(\alpha_r x_n))$, using Proposition 5.2.1. By this identification, additive polynomials are exactly the same as the non-classical polynomials introduced by Tao and Ziegler [164]. As a consequence, we have the following:

Lemma 5.2.4 (Lemma 1.7 of [164]). *$P : \mathbb{K}^n \to \mathbb{T}$ is a polynomial of additive degree d if and only if it can be written in the form:*

$$P(x_1, \ldots, x_n) = \alpha + \sum_{k \geq 0} \sum_{\substack{0 \leq d_{i,j} < p \ \forall i \in [n], j \in [r]: \\ 0 < \sum_{i=1}^{n} \sum_{j=1}^{r} d_{i,j} \leq d - k(p-1)}} \frac{c_{d_{1,1},\ldots,d_{n,r},k} \prod_{i=1}^{n} \prod_{j=1}^{r} |\mathrm{Tr}(\alpha_j x_i)|^{d_{i,j}}}{p^{k+1}} \pmod 1$$

where $\alpha \in \mathbb{T}$ and $c_{d_{1,1},\ldots,d_{n,r}} \in \{0, 1, \ldots, p-1\}$ are uniquely determined. The maximum k for which there is a nonzero $c_{d_{1,1},\ldots,d_{n,r},k}$ is the depth *of P. Note that $depth(P) \leq \left\lfloor \frac{d-1}{p-1} \right\rfloor$ and that P takes on at most $p^{depth(P)+1}$ distinct values.*

For a function $f : \mathbb{K}^n \to \mathbb{C}$, define the *multiplicative derivative in direction* $h \in \mathbb{K}^n$ to be be

$$\Delta_h f(x) = f(x+h) \cdot \overline{f(x)}.$$

5.2.3 Locally Characterized Properties

As described in the introduction, by a locally characterized property, we informally mean a property for which non-membership can be certified by a finite sized witness. Specifically for affine-invariant properties, we define:

Definition 5.2.5 (Locally characterized properties).

- *An* induced affine constraint of size m on ℓ variables *is a pair (A, σ) where A is an affine constraint of size m on ℓ variables and $\sigma \in [R]^m$.*
- *Given such an induced affine constraint (A, σ), a function $f : \mathbb{K}^n \to [R]$ is said to be* (A, σ)-free *if there exist no $x_1, \ldots, x_\ell \in \mathbb{K}^n$ such that $(f(L_1(x_1, \ldots, x_\ell)), \ldots, f(L_m(x_1, \ldots, x_\ell))) = \sigma$. On the other hand, if such $x_1, \ldots, x_\ell$ exist, we say that f* induces (A, σ) at $x_1, \ldots, x_\ell$.

- *Given a (possibly infinite) collection* $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^i, \sigma^i), \ldots\}$ *of induced affine constraints, a function* $f : \mathbb{K}^n \to [R]$ *is said to be* $\mathcal{A}$-free *if it is* (A^i, σ^i)*-free for every* $i \geq 1$*. The size of* $\mathcal{A}$ *is the size of the largest induced affine constraint in* $\mathcal{A}$*.*

- *Additionally,* $\mathcal{A} = \{(A^1, \sigma^1), (A^2, \sigma^2), \ldots, (A^K, \sigma^K)\}$ *is a* W*-light affine system if for some invertible matrix* M*, we have for all* i*,* $MA^i = B^i$ *where* $\mathrm{wt}(B^i) \leq W$*.*

- *A property* $\mathcal{P} \subseteq \{\mathbb{K}^n \to [R]\}$ *is said to be* K, W-lightly locally characterized *if it is equivalent to* $\mathcal{A}$*-freeness for some* W*-light affine system* $\mathcal{A}$ *whose size is* $\leq K$*.*

We recall that Kaufman and Ron [109] show that:

Theorem 5.2.6 ([109])**.** *The property* $\mathcal{P}_d = \{P : \mathbb{K}^n \to \mathbb{K} : \deg(P) \leq d\}$ *is* $q^{\lceil (d+1)/(q-q/p) \rceil}, d$*-lightly locally characterized.*

5.2.4 Factors and Rank

Next, we define a polynomial factor which forms the basis for much of higher order Fourier analysis.

Definition 5.2.7 (Factor)**.** *A* polynomial factor $\mathcal{B}$ *is a sequence of additive polynomials* $P_1, \ldots, P_C : \mathbb{K}^n \to \mathbb{T}$*. We also identify it with the function* $\mathcal{B} : \mathbb{K}^n \to \mathbb{T}^C$ *mapping* x *to* $(P_1(x), \ldots, P_C(x))$*. An* atom *of* $\mathcal{B}$ *is a preimage* $\mathcal{B}^{-1}(y)$ *for some* $y \in \mathbb{T}^C$*. When there is no ambiguity, we will in fact abuse notation and identify an atom of* $\mathcal{B}$ *with the common value* $\mathcal{B}(x)$ *of all* x *in the atom.*

The partition induced by $\mathcal{B}$ *is the partition of* $\mathbb{K}^n$ *given by* $\left\{\mathcal{B}^{-1}(y) : y \in \mathbb{T}^C\right\}$. *The* complexity *of* $\mathcal{B}$*, denoted* $|\mathcal{B}|$*, is the number of defining polynomials* C*. The* order *of* $\mathcal{B}$*, denoted* $\|\mathcal{B}\|$*, is the total number of atoms in* $\mathcal{B}$*. The* degree *of* $\mathcal{B}$ *is the maximum additive degree among its defining polynomials* $P_1, \ldots, P_C$.

Note that due to Lemma 5.2.4, if $\mathcal{B}$ is defined by polynomials $P_1, \ldots, P_C$,

$$\|\mathcal{B}\| = \prod_{i=1}^{C} p^{\deg(P_i)+1}$$

Definition 5.2.8 (Rank)**.** *Let* $d \in \mathbb{N}$ *and* $P : \mathbb{K}^n \to \mathbb{T}$*. Then* $\operatorname{rank}_d(P)$ *is defined as the smallest integer* k *such that there exist functions* $P_1, \ldots, P_k : \mathbb{K}^n \to \mathbb{T}$ *of additive degree* $\leq d-1$ *and a function* $\Gamma : \mathbb{T}^k \to \mathbb{T}$ *such that* $P(x) = \Gamma(P_1(x), \ldots, P_k(x))$*. If* $d = 1$*, then the rank is* 0 *if* P *is a constant function and is* ∞ *otherwise. If* P *is a polynomial of additive degree* d*, then* $\operatorname{rank}(P) = \operatorname{rank}_d(P)$.

Definition 5.2.9 (Rank and Regularity of Polynomial Factor)**.** *Let* $\mathcal{B}$ *be a polynomial factor defined by the sequence* $P_1, \ldots, P_c : \mathbb{K}^n \to \mathbb{T}$ *with respective depths* $k_1, \ldots, k_c$*. Then, the rank of* $\mathcal{B}$ *is* $\min_{(a_1,\ldots,a_c)} \operatorname{rank}(\sum_{i=1}^{c} a_i P_i)$ *where the minimum is over* $(a_1, \ldots, a_c) \in \mathbb{Z}^c$ *such that* $(a_1 \mod p^{k_1+1}, \ldots, a_c \mod p^{k_c+1}) \neq (0, \ldots, 0)$.

Given a polynomial factor $\mathcal{B}$ *and a non decreasing function* $r : \mathbb{Z}^+ \to \mathbb{Z}^+$*,* $\mathcal{B}$ *is* r*-regular if* $\mathcal{B}$ *is of rank at least* $r(|\mathcal{B}|)$.

Definition 5.2.10 (Semantic and Syntactic refinement)**.** *Let* $\mathcal{B}$ *and* $\mathcal{B}'$ *be polynomial factors. A factor* $\mathcal{B}'$ *is a* syntactic refinement *of* $\mathcal{B}$*, denoted by* $\mathcal{B}' \succeq_{syn} \mathcal{B}$ *if the set of polynomials defining* $\mathcal{B}$ *is a subset of the set of polynomials defining* $\mathcal{B}'$*. It is a* semantic refinement*, denoted by* $\mathcal{B}' \succeq_{sem} \mathcal{B}$ *if for every* $x, y \in \mathbb{K}^n$*,* $\mathcal{B}'(x) = \mathcal{B}'(y)$ *implies* $\mathcal{B}(x) = \mathcal{B}(y)$*. Clearly, a syntactic refinement is also a semantic refinement.*

Our next lemma is the workhorse that allows us to convert any factor into a regular one.

Lemma 5.2.11 (Polynomial Regularity Lemma)**.** *Let $r : \mathbb{Z}^+ \to \mathbb{Z}^+$ be a non-decreasing function and $d > 0$ be an integer. Then, there is a function $C^{(r,d)}_{5.2.11} : \mathbb{Z}^+ \to \mathbb{Z}^+$ such that the following is true. Suppose $\mathcal{B}$ is a factor defined by polynomials $P_1, \ldots, P_C : \mathbb{K}^n \to \mathbb{T}$ of additive degree at most d. Then, there is an r-regular factor $\mathcal{B}'$ consisting of polynomials $Q_1, \ldots, Q_{C'} : \mathbb{K}^n \to \mathbb{T}$ of additive degree $\leq d$ such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$ and $C' \leq C^{(r,d)}_{5.2.11}(C)$.*

Moreover, if $\mathcal{B}$ is itself a refinement of some polynomial factor $\hat{\mathcal{B}}$ that has rank $> (r(C') + C')$, then additionally $\mathcal{B}'$ will be a syntactic refinement of $\hat{\mathcal{B}}$.

Proof. Follows directly from Lemma 2.18 of [27] by identifying $\mathbb{K}^n$ with $\mathbb{F}^{rn}$. □

In fact, the regularization process of Theorem 5.2.11 can be implemented in time $O(n^{d+1})$ [30].

5.2.5 Gowers norm and the inverse theorem

Definition 5.2.12. *The* bias *of a function $f : \mathbb{K}^n \to \mathbb{C}$ is defined as* $\mathrm{bias}(f) = |\mathbb{E}_{x \in \mathbb{K}^n} f(x)|$. *For $P : \mathbb{K}^n \to \mathbb{T}$, we use* $\mathrm{bias}(P)$ *to denote* $\mathrm{bias}(e(P))$.

The *Gowers norm* of a function measures the bias of its iterated derivative. Precisely:

Definition 5.2.13 (Gowers norm)**.** *Given a function $f : \mathbb{K}^n \to \mathbb{C}$ and an integer $d \geq 1$, the* Gowers norm of order d *for f is given by*

$$\|f\|_{U^d} = \left|\mathbb{E}_{h_1,\ldots,h_d,x \in \mathbb{K}^n} \left[(\Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_d} f)(x)\right]\right|^{1/2^d}.$$

If $P : \mathbb{K}^n \to \mathbb{T}$, $\|P\|_{U^d}$ *denotes* $\|e(P)\|_{U^d}$.

Note that as $\|f\|_{U^1} = \text{bias}(f)$ the Gowers norm of order 1 is only a semi-norm. However for $d > 1$, it is not difficult to show that $\|\cdot\|_{U^d}$ is indeed a norm.

There is a tight connection between additive polynomials and Gowers norms. In one direction, it is a straightforward consequence of the monotonicity of the Gowers norm ($\|f\|_{U^d} \leq \|f\|_{U^{d+1}}$) and invariance of the Gowers norm with respect to modulation by lower degree polynomials ($\|f\|_{U^{d+1}} = \|f \cdot e(P)\|_{U^{d+1}}$ for polynomials P of additive degree $\leq d$) that if $f : \mathbb{K}^n \to \mathbb{C}$ is δ-*correlated* with a polynomial P of additive degree $\leq d$, meaning

$$|\mathbb{E}_x f(x)e(-P(x))| \geq \delta$$

for some $\delta > 0$, then

$$\|f\|_{U^{d+1}} \geq \delta.$$

In the other direction, we have the following "Inverse theorem for the Gowers norm".

Theorem 5.2.14 (Theorem 1.11 of [164])**.** *Suppose* $\delta > 0$ *and* $d \geq 1$ *is an integer. There exists an* $\varepsilon = \varepsilon_{5.2.14}(\delta, d)$ *such that the following holds. For every function* $f : \mathbb{K}^n \to \mathbb{C}$ *with* $\|f\|_\infty \leq 1$ *and* $\|f\|_{U^{d+1}} \geq \delta$*, there exists a polynomial* $P : \mathbb{K}^n \to \mathbb{T}$ *of additive degree* $\leq d$ *that is* ε*-correlated with* f*, meaning*

$$|\mathbb{E}_{x \in \mathbb{K}^n} f(x)e(-P(x))| \geq \varepsilon.$$

We can be more explicit when $f = e(P)$ for an additive polynomial P.

Theorem 5.2.15 (Theorem 1.20 of [164]). *Suppose $\delta > 0$ and $d \geq 1$ is an integer. There exists an $r = r_{5.2.15}(\delta, d)$ such that the following holds. If a polynomial $P : \mathbb{K}^n \to \mathbb{T}$ with additive degree d satisfies $\|P\|_{U^d} \geq \delta$, then* $\mathrm{rank}(P) \leq r$.

5.3 New Tools

5.3.1 Equidistribution of regular factors

Our results in this section imply that a regular polynomial factor is "as random as possible", subject to the additive degree and depth bounds of its defining polynomials. Let us start with the following simple observation.

Lemma 5.3.1. *Given $\varepsilon > 0$, let $\mathcal{B}$ be a polynomial factor of degree $d > 0$, complexity C and rank $r_{5.3.1}(d, \varepsilon)$, defined by a sequence of additive polynomials $P_1, \ldots, P_C : \mathbb{K}^n \to \mathbb{T}$ having respective depths $k_1, \ldots, k_C$. Suppose $\alpha = (\alpha_1, \ldots, \alpha_C) \in \mathbb{U}_{k_1+1} \times \cdots \times \mathbb{U}_{k_C+1}$. Then:*

$$\mathbf{Pr}_x[\mathcal{B}(x) = \alpha] = \frac{1}{\|\mathcal{B}\|} \pm \varepsilon.$$

Proof. This is standard. See for example Lemma 3.2 of [27]. □

In our applications though, we will often need not just $\mathcal{B}(x)$ to be nearly uniformly distributed but the tuple $(\mathcal{B}(x) : x \in H)$ for a set $H \subseteq \mathbb{K}^n$ to be nearly uniformly distributed. In particular, we consider the case when H is an affine subspace of $\mathbb{K}^n$. The following lemma is key.

Lemma 5.3.2 (Near orthogonality). *Suppose $\mathcal{B}$ is a polynomial factor of degree d and rank $\geq r^{(5.2.15)}(d, \delta)$, defined by the sequence of additive polynomials $P_1, \ldots, P_c$:*

$\mathbb{K}^n \to \mathbb{T}$. *Let* $A = (L_1, \ldots, L_m)$ *be an affine constraint on* ℓ *variables. Let* $\Lambda = (\lambda_{ij})_{i\in[c],j\in[m]}$ *be a tuple of integers. Define:*

$$P_\Lambda(x_1, \ldots, x_k) = \sum_{i\in[c],j\in[m]} \lambda_{ij} P_i(L_j(x_1, \ldots, x_\ell)).$$

Then one of the following is true.

1. *For every* $i \in [c]$*, it holds that* $\sum_{j\in[m]} \lambda_{ij} Q_i(L_j(\cdot)) \equiv 0$ *for all polynomials* $Q_i : \mathbb{K}^n \to \mathbb{T}$ *with the same additive degree and depth as* P_i*. Clearly, this implies* $P_\Lambda \equiv 0$*.*

2. $P_\Lambda \not\equiv 0$*. Moreover,* $\mathrm{bias}(P_\Lambda) \leq \delta$*.*

Proof. For $j \in [m]$, let $(w_{j,1}, \ldots, w_{j,\ell}) \in \mathbb{K}^\ell$ denote the affine form given by L_j. Note that $w_{j,1} = 1$. For $i > 2$, let $|w_{j,i}| = \sum_{k=1}^{r} |\mathrm{Tr}(\alpha_k w_{j,i})|$, and let $|L_j| = \sum_{i=2}^{\ell} |w_{j,i}|$.

For each i, we do the following. If for some j, we have[3] $|L_j| > \deg(\lambda_{i,j} P_i)$, $\lambda_{i,j} \neq 0$, then using Claim 5.2.1, $L_j(x_1, \ldots, x_\ell) = x_1 + \sum_{i=2}^{\ell} \left(\sum_{k=1}^{r} u_{i,k} \cdot \alpha_k\right) x_i$ where each $u_{i,k} \in [0, p-1]$ and $\sum_{i,k} u_{i,k} > \deg(\lambda_{i,j} P_i)$. Using Equation 5.1, we can replace $\lambda_{i,j} P_i(L_j)$ by a $\mathbb{Z}$ linear combination of $P_i(L_{j'})$ where $L_{j'} \preceq L_j$ until no such j exists. Suppose the new coefficients are denoted by $(\lambda'_{i,j})$. If the $\lambda'_{i,j}$ are all zero, then for every $i \in [c]$ individually, $\sum_{j\in[m]} P_i(L_j(x_1, \ldots, x_\ell)) \equiv 0$. Indeed, $\sum_{j\in[m]} Q_i(L_j(x_1, \ldots, x_\ell)) \equiv 0$ for any Q_i with the same additive degree and depth, as the transformation from $\lambda_{i,j}$ to $\lambda'_{i,j}$ did not use any other information about P_i.

[3]Here, $\deg(\cdot)$ refers to the additive degree.

Else some $\lambda'_{i,j} \neq 0$. Also, $|L_j| \leq \deg(\lambda'_{i,j} P_i)$. Then we show the second part of the lemma, that is $|\mathbb{E}[e(P_\Lambda(x_1, \ldots, x_k)]| \leq \delta$.

Suppose without loss of generality that the following is true.

- $\lambda'_{i,1} \neq 0$ for some $i \in [C]$.
- L_1 is maximal in the sense that for every $j \neq 1$, either $\lambda'_{i,j} = 0$ for all $i \in [C]$ or $|w_{j,s}| < |w_{1,s}|$ for some $s \in [\ell]$.

For $\alpha = (\alpha_1, \ldots, \alpha_\ell) \in \mathbb{K}^\ell$ and $y \in \mathbb{K}^n$ and $P : \mathbb{K}^n \to \mathbb{T}$, define

$$\overline{D}_{\alpha,y} P(x_1, \ldots, x_\ell) = P(x_1 + \alpha_1 y, \ldots, x_\ell + \alpha_\ell y) - P(x_1, \ldots, x_\ell).$$

Then

$$\overline{D}_{\alpha,y}(P_i \circ L_j)(x_1, \ldots, x_\ell) = (D_{L_j(\alpha)y} P_i)(L_j(x_1, \ldots, x_\ell)).$$

Let $\Delta = |L_1|$. Define $\alpha_1, \ldots, \alpha_\Delta$ be the set of vectors of the form $(-w, 0, \ldots, 1, 0, \ldots, 0)$ where 1 is in the ith coordinate for $i \in [2, \ell]$ and $0 \leq w < |w_{1,i}|$. Note that $\langle L_1, \alpha_k \rangle \neq 0$ for $k \in [\Delta]$ but for any $j > 1$ there exists some $k \in [\Delta]$ such that $\langle L_j, \alpha_k \rangle = 0$. Thus,

$$\mathbb{E}_{y_1,\ldots,y_\Delta,x_1,\ldots,x_\ell} \left[e\left((\overline{D}_{\alpha_\Delta, y_\Delta} \ldots \overline{D}_{\alpha_1 . y_1} P_\Lambda)(x_1, \ldots, x_\ell) \right) \right] = \left\| \sum_{i=1}^{C} \lambda'_{i,1} P_i \right\|_{U^\Delta}^{2^\Delta}.$$

The rest of the analysis is same as Theorem 3.3 in [27] and we skip it here. □

We can now use Lemma 6.4.10 to prove our result on equidistribution of regular factors over affine subspaces of $\mathbb{K}^n$.

Theorem 5.3.3. *Let $\varepsilon > 0$ Let $\mathcal{B}$ be a polynomial factor defined by polynomials $P_1, \ldots, P_c : \mathbb{K}^n \to \mathbb{T}$ with respective additive degrees $d_1, \ldots, d_c \in \mathbb{Z}^+$ and depths $k_1, \ldots, k_c \in \mathbb{Z}^{\geq 0}$. Suppose $\mathcal{B}$ has rank at least $r^{(5.2.15)}(d, \varepsilon)$ where $d = \max(d_1, \ldots, d_c)$. Let $A = (L_1, \ldots, L_m)$ be an affine constraint. For every $i \in [c]$, define Λ_i to be the set of tuples $(\lambda_1, \ldots, \lambda_m) \in [0, p^{k_i+1}-1]$ such that $\sum_{j=1}^m \lambda_j Q_i(L_j(\cdot)) \equiv 0$ for all polynomials Q_i with the same additive degree and depth as P_i.*

Consider $(\alpha_{i,j} : i \in [c], j \in [m]) \in \mathbb{T}^{cm}$ such that for every $i \in [c]$ and for every $(\lambda_1, \ldots, \lambda_m) \in \Lambda_i$, $\sum_{j=1}^m \lambda_j \alpha_{i,j} = 0$. Then:

$$\mathbf{Pr}_{x_1,\ldots,x_\ell \in \mathbb{K}^n}[\mathcal{B}(L_j(x_1, \ldots, x_\ell)) = (\alpha_{1,j}, \ldots, \alpha_{c,j}) \ \forall j \in [m]] = \frac{\prod_{i=1}^c |\Lambda_i|}{\|\mathcal{B}\|^m} \pm \varepsilon$$

Proof.

$$\begin{aligned}
&\mathbf{Pr}_{x_1,\ldots,x_\ell \in \mathbb{K}^n}[\mathcal{B}(L_j(x_1, \ldots, x_\ell)) = (\alpha_{1,j}, \ldots, \alpha_{c,j}) \ \forall j \in [m]] \\
&= \mathbb{E}_{x_1,\ldots,x_\ell} \left[\prod_{i,j} \frac{1}{p^{k_i+1}} \sum_{\lambda_{i,j}=0}^{p^{k_i+1}-1} e(\lambda_{i,j}(P_i(L_j(x_1, \ldots, x_\ell)) - \alpha_{i,j})) \right] \\
&= \left(\prod_i p^{-(k_i+1)} \right)^m \sum_{\substack{(\lambda_{i,j}) \\ \in \prod_{i,j}[0, p^{k_i+1}-1]}} e\left(-\sum_{i,j} \lambda_{i,j}\alpha_{i,j} \right) \mathbb{E}\left[e\left(\sum_{i,j} \lambda_{i,j} P_i(L_j(x_1, \ldots, x_\ell)) \right) \right] \\
&= p^{-m\sum_{i=1}^c (k_i+1)} \cdot \left(\prod_{i=1}^c |\Lambda_i| \ \pm \ \varepsilon p^{m\sum_{i=1}^c (k_i+1)} \right)
\end{aligned}$$

The last line is due to the observation that from Lemma 6.4.10, $\sum_{i=1}^c \sum_{j=1}^m \lambda_{i,j} P_i(L_j(x_1, \ldots, x_\ell)) \equiv 0$ if and only if for every $i \in [c]$, $(\lambda_{i,1}, \ldots, \lambda_{i,m}) \in \Lambda_i \pmod{p^{k_i+1}}$. So, $\sum_{i,j} \lambda_{i,j} P_i(L_j(\cdot))$ is identically 0 for $\prod_i |\Lambda_i|$ many tuples $(\lambda_{i,j})$ and for those tupes, $\sum_{i,j} \lambda_{i,j}\alpha_{i,j} = 0$ also.

□

5.3.2 Preservation of Locally Characterized Properties

Theorem 5.3.4. *Let $\mathcal{P} \subset \{\mathbb{K}^n \to \mathbb{K}\}$ be a K, W-lightly locally characterized property. For an integer d, suppose $P_1, \ldots, P_c : \mathbb{K}^n \to \mathbb{T}$ are polynomials of additive degree $\leq d$, forming a factor of rank $> r_{5.3.4}(d, K)$, and $\Gamma : \mathbb{T}^c \to \mathbb{K}$ is a function such that $F : \mathbb{K}^n \to \mathbb{K}$ defined by $F(x) = \Gamma(P_1(x), \ldots, P_c(x))$ satisfies $\mathcal{P}$.*

For every collection of additive polynomials $Q_1, \ldots, Q_c : \mathbb{K}^n \to \mathbb{T}$ with $\deg(Q_i) \leq \deg(P_i)$ and $\operatorname{depth}(Q_i) \leq \operatorname{depth}(P_i)$ for all $i \in [c]$, if $G : \mathbb{K}^n \to \mathbb{K}$ is defined by $G(x) = \Gamma(Q_1(x), \ldots, Q_c(x))$, then $G \in \mathcal{P}$ too.

Proof. For the sake of contradiction, suppose $G \notin \mathcal{P}$. Then, by definition, for an affine constraint consisting of K linear forms $L_1, \ldots, L_K$, there exist $x_1, \ldots, x_\ell$ such that $(G(L_1(x_1, \ldots, x_\ell)), \ldots, G(L_K(x_1, \ldots, x_\ell)))$ which form a witness to $G \notin \mathcal{P}$. In other words, there exists $x, y_1, \ldots, y_K \in \mathbb{K}^n$ such that the tuple $B = (Q_i(L_j(x_1, \ldots, x_\ell)) : j \in [K], i \in [c]) \in \mathbb{T}^{cK}$ is a proof of the fact that $G \notin \mathcal{P}$.

Now we argue that there exist $x'_1, \ldots, x'_\ell$ such that $(P_i(L_j(x'_1, \ldots, x'_\ell)) : i \in [c], j \in [K])$ equals B, thus showing that $F \notin P$, a contradiction. Notice that B satisfies the conditions required of α in Theorem 5.3.3. So by Theorem 5.3.3,

$$\mathbf{Pr}_{x'_1, \ldots, x'_\ell}\left[(P_i(L_j(x'_1, \ldots, x'_\ell) : i \in [c], j \in [K]) = B\right] > 0$$

if the rank of the factor formed by $P_1, \ldots, P_c$ is more than $r^{(5.2.15)}\left(d, \frac{1}{2\|\mathcal{B}\|^K}\right)$, where $\|\mathcal{B}\| = p^{\sum_{i=1}^{c}(\operatorname{depth}(P_i)+1)}$. □

In our applications, we will use Theorem 5.3.4 for the property of having bounded degree, which is lightly locally characterized by Theorem 5.2.6.

5.4 List decoding of RM codes

We state the following corollary which we need in the proof to follow. We only state a special case of it which is enough.

Corollary 5.4.1 (Corollary 3.3 of [35]). *Let $g : K \to K$, $\varepsilon > 0$. Then there exist $c \leq 1/\varepsilon^2$ functions $h_1, h_2, \ldots, h_c \in \mathrm{RM}_{\mathbb{K}}(n, d)$ such that for every $f \in \mathrm{RM}_{\mathbb{K}}(n, d)$, there is a function $\Gamma_f : \mathbb{K}^c \to \mathbb{K}$ such that*

$$\mathbf{Pr}_x[\Gamma_f(h_1(x), \ldots, h_c(x)) = f(x)] \geq \mathbf{Pr}_x[g(x) = f(x)] - \varepsilon.$$

Theorem 5.1.1 (Restated). Let $\mathbb{K} = \mathbb{F}_q$ be an arbitrary finite field. Let $\varepsilon > 0$ and $d, n \in \mathbb{N}$. Then,

$$\ell_{\mathbb{K}}(d, n, \delta_{\mathbb{K}}(d) - \varepsilon) \leq c_{q,d,\varepsilon}.$$

Proof. We follow the proof structure in [35]. Let $g : \mathbb{K}^n \to \mathbb{K}$ be a received word. Apply Corollary 5.4.1 with approximation parameter $\varepsilon/2$ gives $\mathcal{H}_0 = \{h_1, \ldots, h_c\} \subseteq \mathrm{RM}_{\mathbb{K}}(n, d)$, $c \leq 4/\varepsilon^2$ such that, for every $f \in \mathrm{RM}_{\mathbb{K}}(n, d)$, there is a function $\Gamma_f : \mathbb{K}^c \to \mathbb{K}$ satisfying

$$\mathbf{Pr}[\Gamma_f(h_1(x), h_2(x), \ldots, h_c(x)) = f(x)] \geq \mathbf{Pr}[g(x) = f(x)] - \varepsilon/2.$$

By Proposition 5.2.1,

$$\mathbf{Pr}[\Gamma'_f(\mathrm{Tr}(\alpha_i h_j(x)) : 1 \leq i \leq r, 1 \leq j \leq c) = F(\mathrm{Tr}(\alpha_i f(x)) : 1 \leq i \leq r)] \geq d/q + \varepsilon/2,$$

where $\Gamma'_f : \mathbb{F}^{rc} \to \mathbb{K}$ and $F : \mathbb{F}^r \to \mathbb{K}$. From here onwards, we identify $\mathbb{F}$ with $\mathbb{U}_1$. Let $\mathcal{H} = \{\mathrm{Tr}(\alpha_i h_j(x)) : 1 \leq i \leq r, 1 \leq j \leq c\}$ and $\mathcal{H}_F = \{\mathrm{Tr}(\alpha_i f(x)) : 1 \leq i \leq r)\}$.

Let $r_1, r_2 : \mathbb{N} \to \mathbb{N}$ be two non decreasing functions to be specified later, and let $C_{r,d}^{(5.2.11)}$ be as given in Lemma 5.2.11. We will require that for all $m \geq 1$,

$$r_1(m) \geq r_2(C_{r_2,d}^{(5.2.11)}(m+1)) + C_{r_2,d}^{(5.2.11)}(m+1) + 1. \tag{5.2}$$

As a first step, we r_1-regularize $\mathcal{H}$ by Lemma 5.2.11. This gives an r_1-regular factor $\mathcal{B}'$ of degree at most d, defined by polynomials $H_1, \ldots, H_{c'} : \mathbb{K}^n \to \mathbb{T}$, $c' \leq C_{r_1,d}^{(5.2.11)}(cr)$ and $\mathrm{rank}(\mathcal{B}') \geq r_1(c')$. We denote $\mathcal{H}' = \{H_1, \ldots, H_{c'}\}$. Let $\mathrm{depth}(H_i) = k_i$ for $i \in [c']$. Let $G_f : \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \to \mathbb{K}$ be defined such that

$$\Gamma_f(h_1(x), \ldots, h_c(x)) = G_f(H_1'(x), \ldots, H_{c'}'(x)).$$

Next, given any polynomial $f : \mathbb{K}^n \to \mathbb{K}$ of degree at most d, we will show that if $\mathbf{Pr}[f(x) \neq g(x)] \leq \delta(d) - \varepsilon$, then f is measurable with respect to $\mathcal{H}'$ and this would upper bound the number of such polynomials by $c'(q, d, \varepsilon)$ independent on n.

Fix such a polynomial f. Call $F_i = \mathrm{Tr}(\alpha_i f)$. Appealing again to Lemma 5.2.11, we r_2-regularize $\mathcal{B}_f := \mathcal{B}' \bigcup \mathcal{H}_F$. We get an r_2-regular factor $\mathcal{B}'' \succeq_{syn} \mathcal{B}'$ defined by the collection $\mathcal{H}'' = \{H_1, \ldots, H_{c'}, H_1', \ldots, H_{c''}'\}$. Note that it is a syntactic refinement of $\mathcal{B}'$ as by our choice of r_1,

$$\mathrm{rank}(\mathcal{B}') \geq r_1(c') \geq r_2(C_{r_2,d}^{(5.2.11)}(c'+1)) + C_{r_2,d}^{(5.2.11)}(c'+1) + 1 \geq r_2(|\mathcal{B}''|) + |\mathcal{B}''| + 1.$$

We will choose r_2 such that for all $m \geq 1$,

$$r_2(m) = \max\left(r_d^{(5.3.1)}\left(\frac{\varepsilon/4}{\left(p^{\lfloor \frac{d-1}{p-1} \rfloor + 1}\right)^m} \right), r_d^{(5.3.4)}(m) \right). \tag{5.3}$$

Let $\mathrm{depth}(H'_i) = k'_i$ for $i \in [c'']$. Let S denote $\otimes_{i=1}^{c'} \mathbb{U}_{k_i+1} \otimes_{j=1}^{c''} \mathbb{U}_{k_j}$. Since each F_i is measurable with respect to $\mathcal{B}''$, there exists $F' : S \to \mathbb{K}$ such that

$$f(x) = F'(H_1(x), \ldots, H_{c'}(x), H'_1(x), \ldots, H'_{c''}(x)).$$

We write G for G_f for brevity. Summing up, we have

$$\mathbf{Pr}[G(H_1(x), H_2(x), \ldots, H_{c'}(x)) = F'(H_1(x), \ldots, H_{c'}(x), H'_1(x), \ldots, H'_{c''}(x))] \geq d/q + \varepsilon/2.$$

We next show that we can have each polynomial in the factor have a disjoint set of inputs. This would simplify the analysis considerably.

Claim 5.4.2. *Let x^i, y^j, $i \in [c'], j \in [c'']$ be pairwise disjoint sets of $n \in \mathbb{N}$ variables each. Let $n' = n(c' + c'')$. Let $\tilde{f} : \mathbb{K}^{n'} \to \mathbb{K}$ and $\tilde{g} : \mathbb{K}^{n'} \to \mathbb{K}$ be defined as*

$$\tilde{f}(x) = F'(H_1(x^1), \ldots, H_{c'}(x^{c'}), H'_1(y^1), \ldots, H'_{c''}(y^{c''}))$$

and

$$\tilde{g}(x) = G(H'_1(x^1), \ldots, H_{c'}(x^{c'})).$$

Then $\deg(\tilde{f}) \leq d$ *and*

$$\left| \mathbf{Pr}_{x \in \mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] - \mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))] \right| \leq \varepsilon/4.$$

Proof. The bound $\deg(\tilde{f}) \leq \deg(f) \leq d$ follows from Lemma 5.3.4 since $r_2(|\mathcal{H}''|) \geq r_d^{(6.4.16)}(|\mathcal{H}''|)$. To establish the bound on $\mathbf{Pr}[\tilde{f} = \tilde{g}]$, for each $s \in S$ let

$$p_1(s) = \mathbf{Pr}_{x \in \mathbb{F}^n}[(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)) = s].$$

Applying Lemma 5.3.1 and since our choice of r_2 satisfies $\text{rank}(\mathcal{H}'') \geq r_d^{(5.3.1)}(\varepsilon/4|S|)$, we have that p_1 is nearly uniform over S,

$$p_1(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

Similarly, let

$$p_2(s) = \mathbf{Pr}_{x^1,\ldots,x^{c'},y^1,\ldots,y^{c''} \in \mathbb{F}^n}[(h'_1(x^1),\ldots,h'_{c'}(x^{c'}),h''_1(y^1),\ldots,h''_{c''}(y^{c''})) = s].$$

Note that the rank of the collection of polynomials $\{h'_1(x^1),\ldots,h'_{c'}(x^{c'}),h''_1(y^1),\ldots,h''_{c''}(y^{c''})\}$ defined over $\mathbb{F}^{n'}$ cannot be lower than that of $\mathcal{H}''$. Applying Lemma 5.3.1 again gives

$$p_2(s) = \frac{1 \pm \varepsilon/4}{|S|}.$$

For $s \in S$, let $s' \in \otimes_{i=1}^{c'} \mathbb{U}_{k_i+1}$ be the restriction of s to first c' coordinates, that is, $s' = (s_1,\ldots,s_{c'})$. Thus

$$\begin{aligned}
\mathbf{Pr}_{x\in\mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] &= \sum_{s\in S} p_2(s) 1_{F(s)=G_f(s')} \\
&= \sum_{s\in S} p_1(s) 1_{F(s)=G_f(s')} \pm \varepsilon/4 \\
&= \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x),h'_2(x),\ldots,h'_c(x))] \pm \varepsilon/4.
\end{aligned}$$

□

So, we obtain that

$$\mathbf{Pr}_{x\in\mathbb{F}^{n'}}[\tilde{f}(x) = \tilde{g}(x)] \geq \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x),\ldots,h'_{c'}(x))] - \varepsilon/4 \geq 1 - \delta(d) + \varepsilon/4.$$

Next, we need the following variant of the Schwartz-Zippel lemma from [35].

Claim 5.4.3. *Let $d, n_1, n_2 \in \mathbb{N}$. Let $f_1 : \mathbb{K}^{n_1+n_2} \to \mathbb{K}$ and $f_2 : \mathbb{K}^{n_1} \to \mathbb{K}$ be such that $\deg(f_1) \leq d$ and*

$$\mathbf{Pr}[f_1(x_1, \ldots, x_{n_1+n_2}) = f_2(x_1, \ldots, x_{n_1})] > 1 - \delta(d)$$

Then, f_1 does not depend on $x_{n_1+1}, \ldots, x_{n_1+n_2}$.

With claim 5.4.3 applied to $f_1 = \tilde{f}, f_2 = \tilde{g}, n_1 = nc', n_2 = nc''$. We obtain that $\tilde{f}$ does not depend on $y^1, \ldots, y^{c''}$. Hence,

$$\tilde{f}(x^1, \ldots, x^{c'}, y^1, \ldots, y^{c''}) = F(H'_1(x^1), \ldots, H'_{c'}(x^{c'}), C_1, \ldots, C_{c''})$$

where $C_j = H''_j(0)$ for $j \in [c'']$. If we substitute $x^1 = \ldots = x^{c'} = x$ we get that

$$f(x) = F(H'_1(x), \ldots, H'_{c'}(x), H''_1(x), \ldots, H''_{c''}(x)) = F(H'_1(x), \ldots, H'_{c'}(x), C_1, \ldots, C_{c''}),$$

which shows that f is measurable with respect to $\mathcal{H}'$, as claimed.

□

5.5 Polynomial decomposition

Definition 5.5.1. *Given $k \in \mathbb{N}$ and $\Delta = (\Delta_1, \ldots, \Delta_k) \in \mathbb{N}^k$ and a function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, a function $P : \mathbb{K}^n \to \mathbb{K}$ is (k, Δ, Γ)-structured if there exist polynomials $P_1, \ldots, P_k : \mathbb{K}^n \to \mathbb{K}$ with $\deg(P_i) \leq \Delta_i$ such that for $x \in \mathbb{K}^n$, we have*

$$P(x) = \Gamma(P_1(x), \ldots, P_k(x)).$$

The polynomials $P_1, \ldots, P_k$ form a (k, Δ, Γ)-decomposition.

The main result we prove is the following.

Theorem 5.5.2. *Let $k \in \mathbb{N}$. For every $\Delta = (\Delta_1, \ldots, \Delta_k) \in \mathbb{N}^k$ and every function $\Gamma : \mathbb{K}^k \to \mathbb{K}$, there is a randomized algorithm A that on input $P : \mathbb{K}^n \to \mathbb{K}$ of degree d, runs in time $\mathrm{poly}_{q,k,\Delta}(n^{d+1})$ and outputs a (k, Δ, Γ)-decomposition of P if one exists while otherwise returning NO.*

We first show that the notion of rank is robust to hyperplane restrictions over nonprime fields. More precisely, we have the following.

Lemma 5.5.3. *Let $P : \mathbb{K}^n \to \mathbb{T}$ be an additive polynomial such that $\mathrm{rank}(P) \geq r$. Let H be a hyperplane in $\mathbb{K}^n$. Then the restriction of P to H has rank at least $r - q$.*

Proof. Without loss of generality, let H be defined by $x_1 = 0$. Let $P' : \mathbb{F}^{n-1} \to \mathbb{F}$ be the restriction of P defined by $P'(y) = P(0y)$. Let $\pi : \mathbb{F}^n \to \mathbb{F}^{n-1}$ be the map $\pi(x_1 x_2 \ldots x_n) = x_2 \ldots x_n$. Let $P'' : \mathbb{F}^n \to \mathbb{F}$ be defined by $P''(x) = P(x) - P' \circ \pi$. Then $P''(x) = 0$ for $x \in H$. For $i \in \mathbb{F}^*$, let $h_i = (i, 0, \ldots, 0)$. Then, for $y \in H$, define $R_j : \mathbb{F}^n \to \mathbb{F}$ by

$$R_j(y) = P''(y + h_j) = (D_{h_j} P'')(y).$$

Note that $\deg(R_j) \leq d - 1$. Now, since $P(x) = P''(x) + P' \circ \pi(x)$, we have

$$P(x) = \Gamma(P' \circ \pi, x_1, \{R_y(x) : y \in \mathbb{F}\}).$$

Now, if $\mathrm{rank}(P') \leq r$, then $\mathrm{rank}(P' \circ \pi) \leq r$ and hence $\mathrm{rank}(P) \leq r + q$. This finishes the proof. □

We now start with the proof of Theorem 5.5.2.

Proof. Let $R_1 : \mathbb{N} \to \mathbb{N}$ be defined as $R_1(m) = R_2(c_{5.2.11}^{(R_1,d)}(m+k)) + c_{5.2.11}^{(R_1,d)}(m+k) + q$ where $R_2 : \mathbb{N} \to \mathbb{N}$ will be specified later.

By Proposition 5.2.1, $P(x) = \sum_i \alpha_i \mathrm{Tr}(\alpha_i P(x))$. Set $f_i(x) = \mathrm{Tr}(\alpha_i P(x))$. Identifying $\mathbb{F}$ with $\mathbb{U}_1$ we treat $f_i : \mathbb{K}^n \to \mathbb{T}$. Regularize $\{f_1, \ldots, f_r\}$ using the algorithm of [30] to find R_1-regular $\mathcal{B} = \{g_1, \ldots, g_C : \mathbb{K}^n \to \mathbb{T}\}$ where $C \leq c_{5.2.11}^{(R_1,d)}(r)$. So, $f_i(x) = G_i(g_1(x), \ldots, g_C(x))$ and $P(x) = \sum_i \alpha_i G_i(g_1(x), \ldots, g_C(x))$. Thus, if $n \leq Cd$, then we are done by a brute force search.

Else, $n > Cd$. For each g_i, pick a monomial m_i with degree $\deg(P_i)$. Then there is $i_0 \in [n]$ such that x_{i_0} does not appear in any g_i. Set $g_i' := g_i | x_{i_0} = 0$. Let $\mathcal{B}'$ be the factor defined by the $g_i's$. Note that $\deg(g_i') = \deg(g_i)$ and $\mathrm{depth}(g_i') = \mathrm{depth}(g_i)$. Also, by Lemma 6.4.18, $\mathcal{B}'$ is $R_1 - q$-regular.

Now, using recursion, we solve the problem on $n-1$ variables. That is, decide if for $P' := P | x_{i_0} = 0$ is (k, Δ, Γ)-structured. If P' is not, then P is not either, so we are done. Else, suppose the algorithm does not output NO.

Say

$$P'(x) = \Gamma(S_1(x), \ldots, S_k(x)) = \Gamma'(\mathrm{Tr}(\alpha_j S_i(x)) : i \in [k], j \in [r]),$$

where

$$\Gamma'(a_{ij} : i \in [k], j \in [r]) = \Gamma(\sum_j \alpha_i a_{ij} : i \in [k]).$$

Note that while $\Gamma : \mathbb{K}^k \to \mathbb{K}$, we have $\Gamma' : \mathbb{F}^{kr} \to \mathbb{K}$. Let $\mathcal{B}_1$ be the factor formed by $\{\mathrm{Tr}(\alpha_j S_i)\}$. Via the algorithm of [30], regularize $\mathcal{B}' \cup \mathcal{B}_1$ using $R_2 : \mathbb{N} \to \mathbb{N}$ and we get a syntactic refinement $\mathcal{B}' \cup \mathcal{B}_1'$ by the choice of R_1. Let $\mathcal{B}_1' = \{s_1', \ldots, s_D'\}$.

where

$$\mathrm{Tr}(\alpha_j S_i) = G_{ij}(g'_i, s'_j : i \in [C], j \in [D]).$$

Choose R_2 large enough such that the map induced by $\mathcal{B}' \cup \mathcal{B}'_1$ is surjective. Now, fix any $\ell \in [r]$. Then,

$$\mathrm{Tr}(\alpha_\ell P') = G_\ell(g'_1, \ldots, g'_C) = F_\ell(G_{ij}(g'_i, s'_j)),$$

where $F_\ell = \mathrm{Tr}(\alpha_\ell \Gamma')$. Thus, for $a_1, \ldots, a_C, b_1, \ldots b_D \in \mathbb{F}$,

$$G_\ell(a_1, \ldots, a_C) = F_\ell(G_{ij}(a_1, \ldots, b_D) : i \in [C], j \in [D]).$$

Substituting, $a_i = g_i(x)$ and $b_j = 0$ we have

$$\mathrm{Tr}(\alpha_\ell P) = G_\ell(g_1, \ldots, g_C) = F_\ell(G_{ij}(g_i, 0)).$$

Now,

$$\mathrm{Tr}(\alpha_\ell P) = \mathrm{Tr}(\alpha_\ell \Gamma(Q_i : i \in [k])),$$

where $Q_i(x) = \sum_{j=1}^{r} \alpha_j G_{ij}(g'_i, \ldots, 0)$.

Since, this is true for all $\ell \in [r]$, we have

$$P(x) = \Gamma(Q_1(x), \ldots, Q_k(x)).$$

where Q_i is defined as above. This finishes the proof. □

Chapter 6

Bias vs low rank of polynomials with applications to list decoding and effective algebraic geometry

6.1 Introduction

Recently, Tao proved an algebraic regularity lemma (Lemma 5 in [162]) which improves upon the Szemeredi regularity lemma [156] in the setting where the graph is definable over a field of large characteristic. The lemma is stated more generally, but in the setting of finite fields it is more straightforward to state. Let $\mathbb{F}$ be a prime field. A set $E \subseteq \mathbb{F}^n$ is a definable set if it is of the form

$$\{(x \in \mathbb{F}^n : p(x) \text{ is true}\},$$

where $p(\cdot)$ is any arbitrary formula involving n variables $x_1, \ldots, x_n$ and a finite number of additional constants $c_1, \ldots, c_m \in \mathbb{F}$ and bound variables $y_1, \ldots, y_l$, as well as the logical and arithmetic operators. For example,

$$V(\mathbb{F}) = \{x \in \mathbb{F}^n : P_1(x) = \ldots, P_m(x) = 0\},$$

is a definable set. Over finite fields, every subset if trivially definable. However, not all have bounded *complexity*. The subset $E \subseteq \mathbb{F}^n$ is definable of complexity at most M if the dimension $n \leq M$ and the length of $p(\cdot)$ is at most M.

The algebraic regularity lemma is informally as follows.

Lemma 6.1.1 (Informal [162])**.** *Let $\mathbb{F}$ be a large prime field. Let V, W be definable sets in $\mathbb{F}^n$ of low complexity and fixed n. Let $E \subseteq V \times W$ be another definable set of low complexity. Then, V and W can be partitioned into small number of V_i and W_j where each of the partitions are also low complexity definable sets, and for all large enough $A \subseteq V_i$ and $B \subseteq W_j$, the edge density of $E \cap (A \times B)$ is similar to the edge density of $E \cap (V_i \times W_j)$.*

The above lemma works for fixed n and improves the parameters of Szemeredi's regularity lemma, in the sense that there are no exceptional pairs and the error in edge density is better.

Let f be a polynomial of degree d in n variables over a finite field $\mathbb{F}$. The polynomial f is said to be unbiased if the distribution of $f(x)$ for a uniform input $x \in \mathbb{F}^n$ is close to the uniform distribution over $\mathbb{F}$, and is called biased otherwise. We say that f has low rank if it can be expressed as a composition of a few lower degree polynomials. The goal is to understand the structure of polynomials that are biased. Green and Tao [80] and Kaufman and Lovett [108] showed over fixed prime fields, that if a fixed degree polynomial is biased, then it has low rank. Such a result lies at the heart of many tools in higher order Fourier analysis. However, the bounds obtained from the above results have very weak dependence (Ackermann-type) on

the field size $|\mathbb{F}|$ and the degree d, and thus are inefficient for large fields. In this work, we extend this to large fields, by proving bounds that are polynomial in the field size $|\mathbb{F}|$.

More precisely, we have the following. Let $\mathbb{F}$ be a prime finite field. Let $\mathcal{P}_d(\mathbb{F}^n)$ denote the family of polynomials $f : \mathbb{F}^n \to \mathbb{F}$ of total degree at most d. Let $e : \mathbb{F} \to \mathbb{C}$ be an additive character, $e(a) = \exp(2\pi i a/|\mathbb{F}|)$.

Theorem 6.1.2. *Let $d, s \in \mathbb{N}$. Let $f \in \mathcal{P}_d(\mathbb{F}^n)$. Suppose that $|\mathbb{E}_{x\in\mathbb{F}^n}[e(f(x))]| \geq |\mathbb{F}|^{-s}$. Then, there exist $g_1, \cdots g_c \in \mathcal{P}_{d-1}(\mathbb{F}^n)$, $c = c^{(6.1.2)}(d, s)$, and $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that $f(x) = \Gamma(g_1(x), \ldots, g_c(x))$.*

Crucially, the rank c is independent of both the field size $|\mathbb{F}|$ and the number of variables n. Moreover, we show (Lemma 6.4.17) that Γ is itself a low degree polynomial: if $\deg(g_i) = d_i$ then

$$\Gamma(z_1, \ldots, z_c) = \sum_{e\in\mathbb{N}^c : \sum d_i e_i \leq d} \alpha_e \prod_{i=1}^{c} z_i^{e_i}.$$

6.1.1 Effective algebraic geometry

Hilbert's strong nullstellensatz establishes a relationship between algebra and geometry and is a fundamental theorem in algebraic geometry. It states the following: given a collection of polynomials, if f vanishes on the set of common zeroes of the polynomials, then some power of f lies in the ideal generated by the collection of polynomials. The area of effective nullstellensatz tries to bound two quantities. First, it bounds what power f should be raised to for the theorem to hold. And

second, it bounds the degrees of the coefficient polynomials when representing a power of f as a member of the ideal. We prove effective versions of these bounds when all polynomials are of fixed degree. Under the regime of fixed degree, we are able to achieve something much stronger, which we highlight shortly. We call this the finite field analogue of the Hilbert nullstellensatz.

Theorem 6.1.3 (Finite field Hilbert Nullstellensatz)**.** *Let $c, d \in \mathbb{N}$. Let $P_1, \ldots P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Assume that $Q(x) = 0$ whenever $P_1(x) = \ldots = P_c(x) = 0$. Then there exist $R_1, \ldots R_c \in \mathcal{P}_D(\mathbb{F}^n)$, $D^{(6.1.3)} = p.O_{d,c}(1)$, such that*

$$Q(x) \equiv \sum_{i=1}^{c} R_i(x) P_i(x).$$

The improvement from the usual nullstellensatz for closed fields here is two-fold:

- There is no exponent in $Q(x)$ which makes this a stronger conclusion.
- $Q(x)$ has to vanish only on the common zeros of the P_i's in the finite field and not its closure which makes this a weaker requirement.

Stated as a Hilbert nullstellensatz result, we have the following.

Theorem 6.1.4 (Effective Hilbert Nullstellensatz)**.** *Let $c, d \in \mathbb{N}$. Let $P_1, \ldots P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Assume that $Q(x) = 0$ whenever $P_1(x) = \ldots = P_c(x) = 0$ for x **in the algebraic closure of** $\mathbb{F}$. Then there exist $R_1, \ldots R_c \in \mathcal{P}_D(\mathbb{F}^n)$, $D^{(6.1.3)} = p.O_{d,c}(1)$ and $\mathbf{r} = \mathbf{1}$, such that*

$$Q(x)^r \equiv \sum_{i=1}^{c} R_i(x) P_i(x).$$

In this work, we focus on the setting of constant d and c. The first effective result in this direction was due to Hermann [95] in 1926 who proved $D = d^{O(2^n)}$ and $r = O_{d,c}(1)$. In 1987, Brownawell [45] and later Kollar [113] in 1988 in breakthrough results, proved a singly exponential bound in n. In fact, for constant c, then achieve $D = O(d^c)$ and $r = O_{d,c}(1)$. Surprisingly, Green and Tao [80] obtained $r = 1$ but $D = O_{d,c,p}(1)$. Note that they have an Ackermann dependence on the field size. However, $r = 1$ immediately leads to a fast ideal membership algorithm. We obtain $D = p.O_{d,c}(1)$ and $r = 1$. Note that we made the dependence on p linear from Ackermann.

A related problem is that of ideal membership. Here the problem is given a collection of polynomials, and a polynomial f, find if f belongs to the ideal generated by the above collection. The challenge is to do this using an efficient algorithm. We prove the following algorithmic result.

Theorem 6.1.5 (Algorithmic Ideal Membership)**.** *Let $c, d \in \mathbb{N}$. Let $P_1, \ldots, P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Let $I = \langle P_1, \ldots, P_c \rangle$. Then we have an algorithm that performs $n^{p.O_{d,c}(1)}$ field operations and decides if $Q \in I$.*

The ideal membership problem is known to be EXPSPACE-hard over the rationals. Over finite fields, the fastest general algorithm was by Buchberger [46] where we formally constructed Gröbner bases in a series of works, with running time d^{2^n}. Green and Tao [80] gave an algorithm with running time $n^{O_{d,c,p}(1)}$. We provide an algorithm with running time $n^{p.O_{d,c}(1)}$. Again, note that the exponent of n was improved from Ackermann to linear in p.

Finally, we come to the problem of counting the number of rational points in a variety. The exact problem of detection of rational points is NP hard. See for example [1, 98, 115, 97, 66, 75].

Given polynomials $f_1, \ldots f_c : \mathbb{F}^n \to \mathbb{F}$, let $V_p(f_1, \ldots, f_c) \subseteq \mathbb{F}^n$ denote the set of common zeroes of f_i's, where the subscript p is to emphasize the interest in rational points.

Lemma 6.1.6 (Rational points in varieties)**.** *Let $c, d, t, u \in \mathbb{N}$. Let $P_1, \ldots, P_c \in \mathcal{P}_d(\mathbb{F}^n)$. There is a randomized algorithm that performs $O_{d,c,t,u}(n^d) + |\mathbb{F}|^{O_{d,c,t}(1)}$ field operations and performs the following with probability $1 - \frac{1}{|\mathbb{F}|^t}$:*

1. *Decide if $V_p(P_1, \ldots, P_c)$ is empty.*
2. *Output an integer N such that $N = (1 \pm |\mathbb{F}|^{-u})|V_p(P_1, \ldots, P_c)|$.*

6.1.2 List Decoding Reed-Muller codes

Recall that the RM code $\mathrm{RM}_{\mathbb{F}}(n, d)$ is defined as follows. The message space consists of degree d polynomials in n variables over $\mathbb{F}$ and the codewords are evaluation of these polynomials on $\mathbb{F}^n$. The distance of two functions $f, g : \mathbb{F}^n \to \mathbb{F}$ is the fraction of points where they disagree,

$$\mathrm{dist}(f, g) := \mathbf{Pr}_x[f(x) \neq g(x)]$$

The minimal distance of a code is the minimal distance of any two distinct codewords. For $\mathrm{RM}_{\mathbb{F}}(n, d)$, this is well understood. When $d < |\mathbb{F}|$ the minimal distance is given by

$$\mathrm{dist}_{\min}(\mathrm{RM}_{\mathbb{F}}(n, d)) = 1 - \frac{d}{|\mathbb{F}|}.$$

More generally, if $d = a(|\mathbb{F}| - 1) + b$ for $0 \leq b \leq |\mathbb{F}| - 1$ then the minimal distance is $|\mathbb{F}|^{-a}(1 - \frac{b}{|\mathbb{F}|})$, but as we focus on large fields, we will always be in the regime of $d < |\mathbb{F}|$.

The list decoding radius of a code is the maximal radius, such that any ball of that radius (centered around an arbitrary function) contains only a few codewords. Let $\mathcal{C} = \mathrm{RM}_{\mathbb{F}}(n, d)$. For $g : \mathbb{F}^n \to \mathbb{F}$, $0 < \rho < 1$ define

$$B_{\mathcal{C}}(g, \rho) := \{f \in \mathcal{P}_d(\mathbb{F}^n) : \mathrm{dist}(f, g) \leq \rho\}.$$

and

$$\ell_{\mathbb{F}}(n, d, \rho) := \max_{g:\mathbb{F}^n \to \mathbb{F}} |B_{\mathcal{C}}(g, \rho)|.$$

The list decoding radius of $\mathcal{C}$ is the maximal radius ρ, up to which $\ell_{\mathbb{F}}(n, d, \rho)$ is "small". In the regime of growing fields, "small" is defined as as polynomial in the field size. It is easy to see that the list decoding radius cannot exceed the minimal distance of the code. The Johnson bound [102] provides a general lower bound for the list decoding radius, which is determined only by the minimal distance of the code. It is known to be tight in general, but it is conjectured not to be tight for special families of codes, for example Reed-Muller codes.

In the regime of constant size fields, it is known that the list decoding radius is in fact equal to the minimal distance of Reed-Muller codes. It was initially proved by Goldreich and Levin [67] and Goldreich, Rubinfield and Sudan [68] for linear polynomials, that is, $d = 1$. Later, Gopalan, Klivans and Zuckerman [74] proved it for the binary field, $\mathbb{F} = \mathbb{F}_2$, and for general fixed prime fields $\mathbb{F}_p$ whenever $(p-1)|d$. They conjectured that is holds for all fixed d, p. Gopalan [73] proved it for $d = 2$.

In Chapter 4 we proved it for all fixed prime fields and all degrees. In this work, we extend this to all prime fields.

Theorem 6.1.7. *Let $d, s \in \mathbb{N}$. There exists $c = c(d, s)$ such that the following holds. For any prime finite field $\mathbb{F}$ with $|\mathbb{F}| > d$ and any $n \in \mathbb{N}$,*

$$\ell_{\mathbb{F}}\left(n, d, 1 - \frac{d}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right) \leq |\mathbb{F}|^c.$$

Moreover, for any $1 \leq e < d$,

$$\ell_{\mathbb{F}}\left(n, d, 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right) \leq |\mathbb{F}|^{c \cdot n^{d-e}}.$$

If $|\mathbb{F}| \leq d$, then the result follows from [35].

6.1.3 Proof Overview

We first present a proof overview for Theorem 6.1.2. The proof is along the lines of Green and Tao [80]. Let $f(x)$ be a polynomial of degree d that is biased, that is $|\mathbb{E}_{x \in \mathbb{F}^n} e(f(x))| \geq |\mathbb{F}|^{-s}$. We first prove that there is a low rank approximation to the given polynomial f. That is, there exist $g_1, \cdots g_c \in \mathcal{P}_{d-1}(\mathbb{F}^n)$, $c = c(d, s, t)$, and $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that

$$\mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) \neq \Gamma(g_1(x), \cdots g_c(x))] \leq |\mathbb{F}|^{-t}.$$

In the regime of fixed finite fields, this was proved by Bogdanov and Viola [39], where the bound c depends polynomially on the underlying parameters, including the error bound, which means that it depends on the field size. Here, we obtain a variant of the lemma, where the bound is independent of the field size. This is crucial in the

next step of the proof, where we show that if the error in approximation is small enough, then it can be converted to an exact computation, if we make the underlying polynomials "random enough" by a regularization process. As this step increases the number of polynomials tremendously, we cannot tolerate any dependence on the field size in the first part of the proof. The proof follows along the lines of [80] with appropriate modifications to tackle the case of growing field size.

The applications in effective algebraic geometry follow by using the principles of regularization, thereby reducing the dimension of the problem to a constant, solving it in constant dimension, and lifting the solution back to the original problem. They are typically straightforward applications of the former result.

The application to list decoding of Reed-Muller codes is more involved and uses the bias vs low rank theorem as one of the building blocks. Given a received function $g : \mathbb{F}^n \to \mathbb{F}$, the first step is to show that it is enough to bound the list size of a subcode of the Reed-Muller code, consisting of only the low rank polynomials. This step is similar to the work of Gopalan [73]. We next show that the list decoding problem for low rank codes can be further reduced to the case where the center g is "nice", concretely, when g is measurable with respect to a small polynomial factor of bounded degree. Unlike the case of fixed finite fields handled in [35], we need to allow a number of potential nice centers for each received word. However, we show that this number is still polynomial in the field size, which allows to keep the number of codewords polynomial in the field size as well. Finally, we prove that the list size around such a nice center is bounded. The last part is similar to the analogous part in the previous work of the authors [35].

6.1.4 Organization

The rest of the paper is as follows. Section 6.2 contains preliminaries. In Section 6.3 we show that any biased polynomial can be approximated by a composition of a small number of lower degree polynomials. In Section 6.4, we show how to convert a good enough approximation to an exact computation. Section 6.5 contains applications in effective algebraic geometry. Section 6.6 contains the application to list decoding of Reed-Muller codes.

6.2 Preliminaries

Let $\mathbb{N}$ denote the set of positive integers. For $n \in \mathbb{N}$, let $[n] := \{1, 2, \ldots, n\}$. We use $y = x \pm \varepsilon$ to denote $y \in [x - \varepsilon, x + \varepsilon]$. For $n \in \mathbb{N}$, and $x, y \in \mathbb{C}^n$, let $\langle x, y \rangle := \sum_{i=1}^{n} x_i \overline{y_i}$ where $\overline{a}$ is the conjugate of a. Let $||x||_2 := \sqrt{\langle x, x \rangle}$.

Fix a prime field $\mathbb{F} = \mathbb{F}_p$. Let $|\cdot| : \mathbb{F} \to \{0, \ldots, p-1\} \subset \mathbb{N}$ be the natural map. Let $e : \mathbb{F} \to \mathbb{C}$ be an additive character, defined as $e(a) := e^{2\pi i a/p}$. Recall that we denote by $\mathcal{P}_d(\mathbb{F}^n)$ the family of polynomials $f : \mathbb{F}^n \to \mathbb{F}$ of total degree at most d. Given a function $f : \mathbb{F}^n \to \mathbb{F}$, its directional derivative in direction $h \in \mathbb{F}^n$ is $D_h f : \mathbb{F}^n \to \mathbb{F}$, given by $D_y f(x) = f(x + y) - f(x)$. Observe that if $f \in \mathcal{P}_d(\mathbb{F}^n)$ then $D_h f \in \mathcal{P}_{d-1}(\mathbb{F}^n)$ for all $h \in \mathbb{F}^n$. For $y_1, \ldots, y_m \in \mathbb{F}^n$ defined the iterative derivative as $D_{y_1,\ldots,y_m} f = D_{y_1} \ldots D_{y_m} f$. In particular, if $f \in \mathcal{P}_d(\mathbb{F}^n)$ and $m > d$ then $D_{y_1,\ldots,y_m} f = 0$.

Let X, Y be finite sets. Define $\Delta(Y) := \{q : Y \to \mathbb{R}_{\geq 0} : \sum_{y \in Y} q(y) = 1\}$ to be the probability simplex on Y. We embed $Y \subset \Delta(Y)$ in the obvious way: $y \in Y$ is mapped to a unit vector e_y with 1 in coordinate y and 0 in all other coordinates. For

a function $f : X \to Y$ let $p(f) : X \to \Delta(Y)$ denote its corresponding embedding, given by $p(f)(x) = e_{f(x)}$. Note that $\Delta(Y)$ is endowed with an inner product, as a subset of $\mathbb{R}^Y$. So, if $f, g : X \to Y$ then

$$\mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = g(x)] = \mathbb{E}_{x \in \mathbb{F}^n}[\langle p(f)(x), p(g)(x) \rangle].$$

6.3 Bias implies low rank approximation

Lemma 6.3.1. *Let $d, s, t \in \mathbb{N}$. Let $f \in \mathcal{P}_d(\mathbb{F}^n)$. Suppose $|\mathbb{E}_{x \in \mathbb{F}^n}[e(f(x))]| \geq |\mathbb{F}|^{-s}$. Then, there exist $g_1, \cdots g_c \in \mathcal{P}_{d-1}(\mathbb{F}^n)$, $c = c(d, s, t) = \binom{d+t+2s+3}{d}$, and $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that*

$$\mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) \neq \Gamma(g_1(x), \cdots g_c(x))] \leq |\mathbb{F}|^{-t}.$$

We prove lemma 6.3.1 in this section. So, fix $f \in \mathcal{P}_d(\mathbb{F}^n)$ and let $\mu = \mathbb{E}_{x \in \mathbb{F}^n}[e(f(x))]$, where we assume $|\mu| \geq |F|^{-s}$. We begin with the following simple claim.

Claim 6.3.2. *For all $x \in \mathbb{F}^n$,*

$$\mu \cdot e(-f(x)) = \mathbb{E}_{y \in \mathbb{F}^n}\left[e(D_y f(x))\right].$$

Proof. $\mathbb{E}_{y \in \mathbb{F}^n}[e(D_y f(x))] = \mathbb{E}_{y \in \mathbb{F}^n}[e(f(x+y))e(-f(x))] = \mathbb{E}_{y \in \mathbb{F}^n}[e(f(y))] \cdot e(-f(x)) = \mu \cdot e(-f(x))$. □

Fix $x \in \mathbb{F}^n$. Pick $z = (z_1, \ldots z_k) \in (\mathbb{F}^n)^k$ uniformly for some k to be specified later. For $a \in \mathbb{F}^k, z \in (\mathbb{F}^n)^k$, we shorthand $a \cdot z = \sum_{i=1}^k a_i z_i \in \mathbb{F}^n$. For $a \in \mathbb{F}^k \setminus \{0\}$, let $W_a(z)$ be the random variable (over the choice of z) defined as

$$W_a(z) := e(D_{a \cdot z} f(x)).$$

For $a \neq 0$, we have

$$\mathbb{E}_z[W_a(z)] = \mathbb{E}_y\left[e(D_y f(x))\right].$$

Also, observe that for distinct $\ell, m \in \mathbb{F}$,

$$|e(\ell) - e(m)| \geq |\mathbb{F}|^{-1}.$$

We have the following.

Claim 6.3.3. *If for $z \in (\mathbb{F}^n)^k$ it holds that*

$$\left|\frac{1}{|\mathbb{F}|^k - 1}\sum_{a\neq 0} W_a(z) - \mathbb{E}_y\left[e(D_y f(x))\right]\right| \leq \frac{1}{2|\mathbb{F}|^{s+1}},$$

then

$$f(x) = \Gamma(D_{a\cdot z} f(x) : a \in \mathbb{F}^k \setminus \{0\})$$

where $\Gamma : \mathbb{F}^{|\mathbb{F}|^k - 1} \to \mathbb{F}$ is some explicit function.

Proof. Since $|e(\ell) - e(m)| \geq |\mathbb{F}|^{-1}$ for $\ell \neq m$ and $|\mu| \geq |\mathbb{F}|^{-s}$, if we define

$$\Gamma(y_1, \ldots y_{|\mathbb{F}|^k - 1}) = \arg\min_{\ell \in \mathbb{F}} \left|\frac{1}{|\mathbb{F}|^k - 1}\sum_{i=1}^{|\mathbb{F}|^k - 1} e(y_i) - e(-\ell)\mu\right|,$$

then by the assumption of the claim,

$$\Gamma(D_{a\cdot z} f(x) : a \in \mathbb{F}^k \setminus \{0\}) = f(x).$$

□

Since the random variables $\{W_a(z) : a \in \mathbb{F}^k \setminus \{0\}\}$ are pairwise independent, we have by Chebychev's inequality that if we choose $k = t + 2s + 3$ then

$$\mathbf{Pr}_{z \in (\mathbb{F}^n)^k} \left[\left| \frac{1}{|\mathbb{F}|^k - 1} \sum_{a \neq 0} W_a(z) - \mathbb{E}_y \left[e(D_y f(x)) \right] \right| \geq \frac{1}{2|\mathbb{F}|^{s+1}} \right] \leq \frac{4|\mathbb{F}|^{2s+2}}{|\mathbb{F}|^k - 1} \leq \frac{1}{|\mathbb{F}|^t}. \tag{6.1}$$

Thus, for all $x \in \mathbb{F}^n$,

$$\mathbf{Pr}_{z \in (\mathbb{F}^n)^k} [\Gamma(D_{a \cdot z} f(x) : a \in \mathbb{F}^k \setminus \{0\}) = f(x)] \geq 1 - |\mathbb{F}|^{-t}.$$

Therefore, by an averaging argument there exists $z \in (\mathbb{F}^n)^k$ for which

$$\mathbf{Pr}_{x \in \mathbb{F}^n} [\Gamma(D_{a \cdot z} f(x) : a \in \mathbb{F}^k \setminus \{0\}) = f(x)] \geq 1 - |\mathbb{F}|^{-t}. \tag{6.2}$$

We now prove our final claim, which shows that we only need a constant number of derivatives in order to approximate f (instead of a number which is polynomial in $|\mathbb{F}|$).

Claim 6.3.4. *Let $\mathcal{B} = \{b \in \mathbb{F}^k : \sum_{j=1}^k |b_j| \leq d\}$. Then for any $a \in \mathbb{F}^k$,*

$$D_{a \cdot z} f(x) = \sum_{b \in \mathcal{B}} \lambda_{a,b} D_{b \cdot z} f(x)$$

for some $\lambda_{a,b} \in \mathbb{F}$.

Proof. Let $|a| = \sum_{i=1}^k |a_i|$. We prove the claim by induction on $|a|$. If $|a| \leq d$ the claim is straightforward, as assume $|a| > d$. As f is a degree d polynomial, we have for any $m > d$ and $y_1, \ldots, y_m \in \mathbb{F}^n$ that

$$D_{y_1} \ldots D_{y_m} f \equiv 0.$$

This translates to

$$\sum_{c\in\{0,1\}^m} (-1)^{\sum c_i} f\left(x + \sum c_i y_i\right) = 0.$$

As the sum of the coefficients is zero, this implies that

$$\sum_{c\in\{0,1\}^m} (-1)^{\sum c_i} D_{c\cdot y} f(x) = 0.$$

Apply this for $m = |a|$ and $y_1, \ldots, y_m$ set to z_1 repeated a_1 times, z_2 repeated a_2 times, up to z_k repeated a_k times. Then we obtain that

$$\sum_{a'\leq a} (-1)^{|a'|} D_{a'\cdot z} f(x) = 0,$$

where the sum is over all $a' \in \mathbb{F}^k$ such that $|a'_i| \leq |a_i|$ for all $1 \leq i \leq k$. We conclude that $D_{a\cdot z} f(x)$ is a linear combination of $D_{a'\cdot z} f(x)$ for $a' \in \mathbb{F}^k$ with $|a'| < |a|$, and apply the induction claim. □

This concludes the proof of Lemma 6.3.1. We can approximate $f(x)$ correctly on $1 - |\mathbb{F}|^{-t}$ fraction of the coordiantes, by a function of $|\mathcal{B}| \leq \binom{d+k}{d}$ polynomials of lower degree, where $k = t + 2s + 3$.

6.4 Bias implies low rank exact computation

The main theorem we prove is the following.

Theorem 6.1.2. Let $d, s \in \mathbb{N}$. Let $f \in \mathcal{P}_d(\mathbb{F}^n)$. Suppose that $|\mathbb{E}_{x\in\mathbb{F}^n}[e(f(x))]| \geq |\mathbb{F}|^{-s}$. Then, there exist $g_1, \cdots g_c \in \mathcal{P}_{d-1}(\mathbb{F}^n)$, $c = c^{(6.1.2)}(d, s)$, and $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that $f(x) = \Gamma(g_1(x), \ldots g_c(x))$.

The proof is by induction on the degree d. But first, we define the notion of regularity followed by some important consequences of Theorem 6.1.2 which are required in the inductive proof of the same and might be of independent interest.

6.4.1 Basic definitions

We recall some of the preliminaries from the previous chapters.

Definition 6.4.1 (Rank)**.** *Let $d \in \mathbb{N}$ and $f : \mathbb{F}^n \to \mathbb{F}$. Then $\mathrm{rank}_d(f)$ is defined as the smallest integer r such that there exist polynomials $h_1, \ldots, h_r : \mathbb{F}^n \to \mathbb{F}$ of degree $\leq d-1$ and a function $\Gamma : \mathbb{F}^r \to \mathbb{F}$ such that $f(x) = \Gamma(h_1(x), \ldots, h_r(x))$. If $d = 1$, then the rank is 0 if f is a constant function and is ∞ otherwise. If f is a polynomial, then $\mathrm{rank}(f) = \mathrm{rank}_d(f)$ where $d = \deg(f)$.*

Definition 6.4.2 (Factor)**.** *Let X be a finite set. Then a factor $\mathcal{B}$ is a partition of the set X. The subsets in the partition are called atoms.*

For finite sets X and Y, recall that $\Delta(Y)$ is the probability simplex over Y, and that we embed $Y \subset \Delta(Y)$ and embed functions $f : X \to Y$ as functions $f : X \to \Delta(Y)$ in the obvious way. For a factor $\mathcal{B}$ of X, a function $f : X \to \Delta(Y)$ is said to be measurable with respect to $\mathcal{B}$ if it is constant on the atoms of $\mathcal{B}$. The average of f over $\mathcal{B}$ is $\mathbb{E}[f|\mathcal{B}] : X \to \Delta(Y)$ defined as

$$\mathbb{E}[f|\mathcal{B}](x) = \mathbb{E}_{y \in \mathcal{B}(x)}[f(y)]$$

where $\mathcal{B}(x)$ is the atom containing x. Clearly, $\mathbb{E}[f|\mathcal{B}]$ is measurable with respect to $\mathcal{B}$.

A collection of functions $h_1, \ldots, h_c : X \to Y$ defines a factor $\mathcal{B}$ whose atoms are $\{x \in X : h_1(x) = y_1, \ldots, h_c(x) = y_c\}$ for every $(y_1, \ldots, y_c) \in Y^c$. We use $\mathcal{B}$ to also denote the map $x \mapsto (h_1(x), \ldots, h_c(x))$. A function f is measurable with respect to a collection of functions if it is measurable with respect to the factor the collection defines.

Definition 6.4.3 (Polynomial Factor)**.** *A polynomial factor $\mathcal{B}$ is a factor defined by a collection of polynomials $\mathcal{H} = \{h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{F}\}$ and the factor is written as $\mathcal{B}_{\mathcal{H}}$. The degree of the factor is the maximum degree of $h \in \mathcal{H}$. With a slight abuse of notation, we would typically identify $\mathcal{H}$ and $\mathcal{B}_{\mathcal{H}}$.*

Let $|\mathcal{B}|$ be the number of polynomials defining the factor. We define $||\mathcal{B}|| := |\mathbb{F}|^c$ to be the number of (possibly empty) atoms.

Definition 6.4.4 (Rank and Regularity of Polynomial Factor)**.** *Let $\mathcal{B}$ be a polynomial factor defined by $h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{F}$. Then, the rank of $\mathcal{B}$ is the least integer r such that there exists $(a_1, \ldots, a_c) \in \mathbb{F}^c$, $(a_1, \ldots, a_c) \neq (0, \ldots, 0)$ for which the linear combination $h(x) := \sum_{i=1}^{c} a_i h_i(x)$ has $\mathrm{rank}_d(h) \leq r$ where $d = \max_i \deg(a_i h_i)$. For a non decreasing function $r : \mathbb{N} \to \mathbb{N}$, a factor $\mathcal{B}$ is r-regular if its rank is at least $r(|\mathcal{B}|)$.*

Definition 6.4.5 (Semantic and Syntactic refinement)**.** *Let $\mathcal{B}$ and $\mathcal{B}'$ be polynomial factors on $\mathbb{F}^n$. A factor $\mathcal{B}'$ is a syntactic refinement of $\mathcal{B}$, denoted by $\mathcal{B}' \succeq_{syn} \mathcal{B}$ if the set of polynomials defining $\mathcal{B}$ is a subset of the set of polynomials defining $\mathcal{B}'$. It is a semantic refinement, denoted by $\mathcal{B}' \succeq_{sem} \mathcal{B}$ if for every $x, y \in \mathbb{F}^n$, $\mathcal{B}'(x) = \mathcal{B}'(y)$ implies $\mathcal{B}(x) = \mathcal{B}(y)$.*

Lemma 6.4.6 (Polynomial Regularity Lemma)**.** *Let $r : \mathbb{N} \to \mathbb{N}$ be a non-decreasing function and $d \in \mathbb{N}$. Then there is a function $C_{r,d}^{(6.4.6)} : \mathbb{N} \to \mathbb{N}$ such that the following is true. Let $\mathcal{B}$ be a factor defined by polynomials $h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{F}$ of degree at most d. Then, there is an r-regular factor $\mathcal{B}'$ defined by polynomials $h'_1, \ldots, h'_{c'} : \mathbb{F}^n \to \mathbb{F}$ of degree at most d such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$ and $c' \leq C_{r,d}^{(6.4.6)}(c)$.*

Moreover if $\mathcal{B} \succeq_{syn} \hat{\mathcal{B}}$ for some polynomial factor $\hat{\mathcal{B}}$ that has rank at least $r(c') + c' + 1$, then $\mathcal{B}' \succeq_{syn} \hat{B}$.

The proof of Lemma 6.4.6 is exactly along the lines of existing proofs in the literature, for example Lemma 2.3 in [80], so we do not repeat it here.

For $(w_1, \ldots, w_k), (w'_1, \ldots, w'_k) \in \mathbb{F}^k$, we write $(w_1, \ldots, w_k) \prec (w'_1, \ldots, w'_k)$ if $|w_i| \leq |w'_i|$ for all $i \in [k]$, where $|\cdot|$ is the canonical map from $\mathbb{F}$ to $\{0, 1, \ldots, p-1\}$.

Definition 6.4.7 (Affine system)**.** *An affine system is a set of linear forms $\{L_1, \ldots, L_m\}$, where each $L_i : \mathbb{F}^k \to \mathbb{F}$ is defined by $L_i(x) = \sum_{j=1}^k w_{i,j} x_j$, which satisfies the following:*

- *$w_{i,1} = 1$ for all $i \in [m]$.*
- *If $L'(x) = \sum_{j=1}^k w'_j x_j$, where $w'_1 = 1$ and $w' \prec w_i$ for some $i \in [m]$, then $w' = w_j$ for some $j \in [m]$.*

6.4.2 Inverse Gowers norm for polynomial phases

Theorem 6.4.8. *Suppose Theorem 6.1.2 is true up to order d. Let $d, s \in \mathbb{N}$ with $d < |\mathbb{F}|$. Let $f \in \mathcal{P}_d(\mathbb{F}^n)$. Suppose $||e(f)||_{U^d} \geq |\mathbb{F}|^{-s}$. Then,* $\mathrm{rank}(f) \leq c^{(6.4.8)}(d, s)$.

Proof. We have

$$|\mathbb{E}_{x,y_1,\ldots,y_d\in\mathbb{F}^n}\left[e\left(D_{y_1,\ldots,y_d}f(x)\right)\right]| = ||e(f)||_{U^d}^{2^d} \geq |\mathbb{F}|^{-s\cdot 2^d}.$$

Let $g : \mathbb{F}^{n(d+1)} \to \mathbb{F}$ be defined as

$$g(x, y_1, \ldots, y_d) := D_{y_1,\ldots,y_d}f(x).$$

By Theorem 6.1.2,

$$\text{rank}(g) \leq c^{(6.1.2)}(d, s \cdot 2^d).$$

By Taylor's theorem, since we assume $d < |\mathbb{F}|$,

$$f(x) = \frac{D_{x,\ldots,x}f(0)}{d!} + h(x),$$

where $h \in \mathcal{P}_{d-1}(\mathbb{F}^n)$. Since, $g(0, x, \ldots, x) \equiv D_{x,\ldots,x}f(0)$, we conclude that $\text{rank}(f) \leq \text{rank}(g) + 1 \leq c^{(6.1.2)}(d, s \cdot 2^d) + 1$. Choosing $c^{(6.4.8)}(d, s)$ large enough such that $c^{(6.4.8)}(d, s) \geq c^{(6.1.2)}(d, s \cdot 2^d) + 1$ finishes the proof. □

6.4.3 Equidistribution of atoms

The next lemma shows that a regular factor has atoms of roughly equal size.

Lemma 6.4.9 (Size of atoms)**.** *Suppose Theorem 6.1.2 is true up to order d. Let $\mathcal{B} = \{h_1, \ldots, h_c\}$ be a polynomial factor of degree at most d. Given $s \in \mathbb{N}$, assume that $\mathcal{B}$ has rank at least $c^{(6.1.2)}(d, s)$. Then for every $b \in \mathbb{F}^c$,*

$$\mathbf{Pr}_{x\in\mathbb{F}^n}[\mathcal{B}(x) = b] = \frac{1}{||\mathcal{B}||} \pm \frac{1}{|\mathbb{F}|^s}.$$

Proof. For any $b \in \mathbb{F}^c$,

$$\begin{aligned}
\mathbf{Pr}[\mathcal{B}(x) = b] &= \frac{1}{|\mathbb{F}|^c} \sum_{a \in \mathbb{F}^c} \mathbb{E}_x \left[e\left(\sum_i a_i (h_i(x) - b_i) \right) \right] \\
&= \frac{1}{|\mathbb{F}|^c} \pm \frac{1}{|\mathbb{F}|^c} \sum_{0 \neq a \in \mathbb{F}^c} \left| \mathbb{E}_x \left[e\left(\sum_i a_i h_i(x) \right) \right] \right| \\
&= \frac{1}{|\mathbb{F}|^c} \pm \frac{1}{|\mathbb{F}|^s}
\end{aligned}$$

The last line follows because of the following. Suppose for some $a \neq 0$, $|\mathbb{E}_x \left[e\left(\sum_i a_i h_i(x) \right) \right]| > \frac{1}{|\mathbb{F}|^s}$, then by Theorem 6.1.2, $\mathrm{rank}(\sum_i a_i h_i) \leq c^{(6.1.2)}(d, s)$. This contradicts the assumption on the rank of $\mathcal{B}$. □

6.4.4 Near orthogonality of affine linear forms

Lemma 6.4.10 (Near orthogonality)**.** *Suppose Theorem 6.1.2 is true up to order d. Let $c, d, p, s, m, k \in \mathbb{N}$. Let $\mathcal{B} = \{h_1, \ldots, h_c\}$ be a polynomial factor of degree at most d. Assume $\mathcal{B}$ has rank at least $r^{(6.4.10)}(d, k, s)$. Let $(L_1, \ldots, L_m)$ be an affine system on k variables. Let $\Lambda = (\lambda_{ij})_{i \in [c], j \in [m]}$ be a tuple of integers. Define*

$$h_\Lambda(x_1, \ldots, x_k) = \sum_{i \in [c], j \in [m]} \lambda_{ij} h_i(L_j(x_1, \ldots, x_k)).$$

Then one of the following is true.

1. *$h_\Lambda \equiv 0$. Moreover, for every $i \in [c]$, it holds that $\sum_{j=1}^m \lambda_{ij} g_i(L_j(\cdot)) \equiv 0$ for all $g_i \in \mathcal{P}_d(\mathbb{F}^n)$.*

2. *$h_\Lambda \not\equiv 0$. Moreover, $|\mathbb{E}[e(h_\Lambda(x_1, \ldots, x_k)]| \leq |\mathbb{F}|^{-s}$.*

Again, the proof is exactly along the lines of Theorem 3.3 in [26] taking care of the dependence on $|\mathbb{F}|$ now, followed by an application of Theorem 6.4.8. As a corollary, we state the above result for the case of parallelepipeds. We will need this in the inductive proof of Theorem 6.1.2.

6.4.5 Equidistribution of parallelepipeds

We first set up some definitions following Section 4 in [80]. Throughout this subsection, let $\mathcal{B} = \{h_1, \ldots, h_c\}$ be a polynomial factor of degree at most d. We assume $\mathcal{B}$ has rank at least $r^{(6.4.8)}(d, s)$. For $i \in [d]$, M_i denotes the number of polynomials in $\mathcal{B}$ of degree exactly equal to i. Let $\Sigma := \otimes_{i\in[d]}\mathbb{F}^{M_i}$.

Definition 6.4.11 (Faces and lower faces)**.** *Let $k \in \mathbb{N}$ and $0 \leq k' \leq k$. A set $F \subseteq \{0,1\}^k$ is called a face of dimension k' if*

$$F = \{b : b_i = \delta_i, i \in I\},$$

where $I \subseteq [k]$, $|I| = k - k'$ and $\delta_i \in \{0,1\}$. If $\delta_i = 0$ for all $i \in I$, the F is a lower face. Thus, it is equivalent to the power set of $[k] \setminus I$.

Definition 6.4.12 (Face vectors and parallelepiped constraints)**.** *Let $i_0 \in [d]$, $j_0 \in [M_{i_0}]$ and $F \subseteq \{0,1\}^k$. Let $r(i_0, j_0, F) \in \Sigma^{\{0,1\}^k}$ indexed as $r(i, j, \omega) = (-1)^{|\omega|}$ if $i = i_0, j = j_0$ and $\omega \in F$ and zero otherwise. This is called a face vector. If F is a lower face, then it corresponds to a lower face vector. If $\dim(F) \geq i_0 + 1$, then it is a relevant face (lower face) vector. A vector $(t(\omega) : \omega \in \{0,1\}^k) \in \Sigma^{\{0,1\}^k}$ satisfies the parallelepiped constraints if it is orthogonal to all the relevant lower face vectors.*

Let $\Sigma_0 \subseteq \Sigma^{\{0,1\}^k}$ be the subspace of vectors satisfying the parallelepiped constraints.

Claim 6.4.13 (Dimension of Σ_0, Lemma 4.4 [80])**.** *Let $d < k$. Then,*

$$\dim(\Sigma_0) = \sum_{i=1}^{d} M_i \sum_{0 \le j \le i} \binom{k}{j}.$$

Lemma 6.4.14 (Equidistribution of parallelepipeds)**.** *Suppose Theorem 6.1.2 is true up to order d. Given $s, d < k \in \mathbb{N}$, let $\mathcal{B}$ be a polynomial factor of rank at least $c^{(6.4.14)}(k, s)$ defined by polynomials $h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{F}$ of degree at most d. For every $t \in \Sigma_0$ and x such that $\mathcal{B}(x) = t(0)$,*

$$\mathbf{Pr}_{y_1,\ldots,y_k}[\mathcal{B}(x + \omega \cdot y) = t(\omega) \ \forall \ \omega \in \{0,1\}^k] = \frac{1}{|\mathbb{F}|^{\sum_{i=1}^{d} M_i \sum_{1 \le j \le i} \binom{k}{j}}} \pm \frac{1}{|\mathbb{F}|^s}.$$

Proof. This immediately follows from the dimension of Σ_0 (Claim 6.4.13) and Lemma 6.4.10 applied to the parallelepiped. □

6.4.6 Proof of Theorem 6.1.2

The proof of Theorem 6.1.2 is by induction on d and follows along the lines of Theorem 1.7 in [80]. We sketch the proof here.

Proof of Theorem 6.1.2. The base case of $d = 1$ is trivial. Indeed, if a linear polynomial $f : \mathbb{F}^n \to \mathbb{F}$ satisfies $|\mathbb{E}[e(f(x)]| \ge |\mathbb{F}|^{-s}$, then by orthogonality of linear polynomials, we have $f(x)$ is a constant and hence has rank 0. Now, suppose the hypothesis is true for degree $d-1$. Let $t \in \mathbb{N}$ depending on d be specified later. We have $|\mathbb{E}[e(f(x))]| \ge |\mathbb{F}|^{-s}$. By Lemma 6.3.1, there exists $\mathcal{B} = \{h_1, \ldots h_c : h_i \in \mathcal{P}_{d-1}(\mathbb{F}^n)\}$, $c = c(d, s, t)$, and $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that

$$\mathbf{Pr}[f(x) \ne \Gamma(h_1(x), \cdots h_c(x))] \le |\mathbb{F}|^{-t}.$$

Let $r : \mathbb{N} \to \mathbb{N}$ be a growth function that depends on d and will be specified later. Regularize $\mathcal{B}$ to an r-regular polynomial factor $\mathcal{B}' = \{h'_1, \ldots, h'_{c'}\}$, $c' \leq C_{r,d}^{(6.4.6)}(c)$. Thus, we have for an appropriate $\Gamma' : \mathbb{F}^{c'} \to \mathbb{F}$ that

$$\mathbf{Pr}[f(x) \neq \Gamma'(h'_1(x), \cdots h'_{c'}(x))] \leq |\mathbb{F}|^{-t}.$$

In the rest of the proof, we prove that f is $\mathcal{B}'$-measurable. This will finish the proof. We will assume that $r(j) \geq c^{(6.4.9)}(d, 2t + j)$ for all $j \in \mathbb{N}$. By Markov's inequality and Lemma 6.4.9, for at least $1 - |\mathbb{F}|^{-t/4}$ fraction of atoms A,

$$\mathbf{Pr}_{x \in A}[f(x) \neq \Gamma'(h'_1(x), \cdots h'_{c'}(x))] \leq |\mathbb{F}|^{-t/4}.$$

The first step is to prove that on such atoms, f is constant. Fix such an atom A and let $A' \subseteq A$ be the set where $f(x) = \Gamma'(h'_1(x), \cdots h'_{c'}(x))$.

Lemma 6.4.15. *Let t be large enough depending on d. Let $x \in A$ be arbitrary. Then there is an $h \in (\mathbb{F}^n)^{d+1}$ such that $x + \omega \cdot h \in A'$ for all $\omega \in \{0,1\}^{d+1} \setminus 0^{d+1}$.*

The proof is exactly as in Lemma 5.2 in [80]. We omit it here. Continuing, since $f \in \mathcal{P}_d(\mathbb{F}^n)$, we have

$$\sum_{\omega \in \{0,1\}^{d+1}} (-1)^{|\omega|} f(x + \omega \cdot h) = 0.$$

Now, by the above lemma, we have $f(x + \omega \cdot h) \equiv c_A$ for $\omega \neq 0$, where c_A is a constant that depends on A. Thus, $f(x) \equiv c_A$.

This finishes the first step. Thus, we have for $1 - |\mathbb{F}|^{-t/4}$ fraction of the atoms A, call them good atoms, $f(x) = c_A$. The final step shows that for any

arbitrary atom A, there are good atoms A_ω, $0 \neq \omega \in \{0,1\}^{d+1}$ such that the vector $t = \mathcal{B}(A_\omega) \in \Sigma^{\{0,1\}^{d+1}}$ satisfies the parallelepiped constraints. It is enough to find one parallelepiped for which $x + \omega \cdot h$ lie in good atoms for $\omega \neq 0$. Indeed, let $x \in A$ be arbitrary. Pick $h_1, \ldots, h_{d+1}$ randomly. The probability that for a fixed $\omega \neq 0$, $x + \omega \cdot h$ lies in a good atom is at least $1 - |\mathbb{F}|^{-t/4} > 1 - 2^{-2d}$ for t large enough. The result now follows by a union bound over $\omega \in \{0,1\}^{d+1}$. □

6.4.7 Some more consequences

Degree preserving lemma.

Lemma 6.4.16 (Degree Preserving Lemma)**.** *Let $c, d, D \in \mathbb{N}$ with $d < |\mathbb{F}|$. Let $\mathcal{B} = \{h_1, \ldots, h_c\}$ be a polynomial factor of degree at most d, and rank at least $r^{(6.4.16)}(c, d, D)$. For $\Gamma : \mathbb{F}^c \to \mathbb{F}$, let $F : \mathbb{F}^n \to \mathbb{F}$ be defined by $F(x) = \Gamma(h_1(x), \ldots, h_c(x))$. Let $\deg(F) = D$. Then, for every set of polynomials $h'_1, \ldots h'_c : \mathbb{F}^n \to \mathbb{F}$ with $\deg(h'_i) \leq \deg(h_i)$ for all $i \in [c]$, if $G : \mathbb{F}^n \to \mathbb{F}$ is defined by $G(x) = \Gamma(h'_1(x), \ldots, h'_c(x))$, we have $\deg(G) \leq D$.*

We omit the proof here as it can be readily adapted from Theorem 4.1 in [26].

Faithful composition.

Lemma 6.4.17 (Faithful composition lemma)**.** *Let $c, d, D \in \mathbb{N}$. Let $\mathcal{B} = \{h_1, \ldots, h_c\}$ be a polynomial factor of degree at most d, and rank at least $r^{(6.4.16)}(c, d, D)$. Let $\Gamma : \mathbb{F}^c \to \mathbb{F}$ be defined by $\Gamma(z) = \sum_{s \in S} a_s \prod_i z_i^{s_i}$, for some $S \subset \mathbb{N}^c$ and where $a_s \neq 0$ for all $s \in S$. Define $F : \mathbb{F}^n \to \mathbb{F}$ by*

$$F(x) = \sum_{s \in S} a_s \prod_{i=1}^{c} h_i(x)^{s_i}.$$

Assume that $\deg(F) = D$. *Then for every* $s \in S$,

$$\sum_{i=1}^{c} s_i \cdot \deg(h_i) \leq D.$$

Proof. Let $d_i = \deg(h_i)$. Define new variables $x' = \{x'_{i,j} : i \in [c], j \in [d_i]\}$ with $x'_{i,j} \in \mathbb{F}$. Define new polynomials $h'_i(x') = \prod_{j=1}^{d_i} x'_{i,j}$, where we note that $h'_1, \ldots, h'_c$ are defined over disjoint sets of variables, and that $\deg(h'_i) = \deg(h_i)$. Define $G(x') = \Gamma(h'_1(x'), \ldots, h'_c(x'))$. Since $\mathcal{B}$ has rank at least $r^{(6.4.16)}(c, d, D)$, we have by Lemma 6.4.16 that $\deg(G) \leq D$. Expanding the definition of Γ we have

$$G(x') = \sum_{s \in S} a_s \prod_{i=1}^{c} \prod_{j=1}^{d_i} (x'_{i,j})^{s_i}.$$

Note that each $s \in S$ corresponds to a unique monomial of degree $\sum_{i=1}^{c} d_i s_i$, and the monomials cannot cancel each other. The lemma follows. □

Hyperplane Restriction. Next, we show that the notion of rank is robust to hyperplane restrictions. More precisely, we have the following.

Lemma 6.4.18. *Let* $f \in \mathcal{P}_d(\mathbb{F}^n)$ *such that* $\mathrm{rank}(P) \geq r$. *Let* H *be a hyperplane in* $\mathbb{F}^n$. *Then the restriction of* f *to* H *has rank at least* $r - d - 1$.

We note that the existing results prove a lower bound of $r - |\mathbb{F}|$, but with a slight modification (which we show below) we are able to prove a lower bound of $r - d - 1$.

Proof. Without loss of generality, let H be defined by $x_1 = 0$. For $x \in \mathbb{F}^n$ let $x' = x_2 \ldots x_n \in \mathbb{F}^{n-1}$ so that $x = (x_1, x')$ and $f|_H(x') = f(0, x')$. Define $f_i : \mathbb{F}^{n-1} \to \mathbb{F}$ by

$$f_i(x') = f(i, x') - f(0, x').$$

Clearly, $f(x) = \Gamma(x_1, f|_H(x), f_1(x'), \ldots, f_{|\mathbb{F}|-1}(x'))$ for some explicit $\Gamma : \mathbb{F}^{|\mathbb{F}|+1} \to \mathbb{F}$. For $v_i = (i, 0, \ldots, 0) \in \mathbb{F}^n$, we have that f_i is the restriction of $D_{v_i} f$ to H, and hence $\deg(f_i) \leq \deg(D_{v_i} f) \leq d - 1$. To conclude the proof, we show that for any $j > d$, $f_j(x')$ can be expressed a linear combination of $\{f_1(x'), \ldots, f_d(x')\}$. This will imply that in fact, $f(x) = \Gamma'(x_1, f|_H, f_1(x'), \ldots, f_d(x'))$ for some $\Gamma' : \mathbb{F}^{d+2} \to \mathbb{F}$ and, since $\deg(f_i) < \deg(f)$ for all i, will show that $\mathrm{rank}(f) \leq \mathrm{rank}(f|_H) + d + 1$.

To conclude the proof, fix $j > d$. We will show that $f_j(x')$ is a linear combination of $\{f_1(x'), \ldots, f_{j-1}(x')\}$, which by induction will show the claim. As $\deg(f) \leq d$ we have

$$\underbrace{D_{x_1} D_{x_1} \ldots D_{x_1}}_{j \text{ times}} f(x) = 0.$$

Writing this explicitly, and restricting to $x = (0, x')$, we obtain that

$$\sum_{i=0}^{j} (-1)^i \binom{j}{i} f(i, x') = 0,$$

which in turn implies that

$$\sum_{i=1}^{j} (-1)^i \binom{j}{i} f_i(x') = 0.$$

Thus, $f_j(x')$ is a linear combination of $\{f_1(x'), \ldots, f_{j-1}(x')\}$, as claimed. □

6.4.8 Algorithmic Aspects

It is easy to see that the existential proof of the main theorem can be made algorithmic.

Lemma 6.4.19. *Let $d, s \in \mathbb{N}$. There is a randomized algorithm that on input $f \in \mathcal{P}_d(\mathbb{F}^n)$ with $|\mathbb{E}[e(f(x))]| \geq |\mathbb{F}|^{-s}$, runs in time $O(|\mathbb{F}|^c \cdot n^d)$ and outputs $g_1, \ldots, g_c \in \mathcal{P}_{d-1}(\mathbb{F}^n)$, $c = c^{(6.1.2)}(d, s)$, and $\Gamma : \mathbb{F}^c \to \mathbb{F}$, such that $f(x) = \Gamma(g_1(x), \ldots g_c(x))$.*

The proof of the above follows similar to Theorem 1.4 in [31]. It can be derandomized using either Viola's generator [169] or Bogdanov's generator [40] for low degree polynomials. To use Bogdanov's generator, one requires the field size to be at least superlogarithmic in n. For details, see the proof of Theorem 1.2 in [24].

6.5 Applications: Effective algebraic geometric bounds over large finite fields

6.5.1 A finite field Hilbert nullstellensatz

We prove the following effective version of a finite field analogue of Hilbert Nullstellensatz.

Theorem 6.1.3. Let $c, d \in \mathbb{N}$. Let $P_1, \ldots P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Assume that $Q(x) = 0$ whenever $P_1(x) = \ldots = P_c(x) = 0$. Then there exist $R_1, \ldots R_c \in \mathcal{P}_D(\mathbb{F}^n)$, $D^{(6.1.3)} = p.O_{d,c}(1)$, such that

$$Q(x) \equiv \sum_{i=1}^{c} R_i(x) P_i(x).$$

Proof. Let $\mathcal{B} = \{P_1, \ldots P_c, Q\}$ be the corresponding polynomial factor. We regularize $\mathcal{B}$ to obtain $\mathcal{B}' = \{S_1, \ldots S_{c'}\}$ with a growth function $r : \mathbb{N} \to \mathbb{N}$ satisfying $r(j) \geq c^{(6.1.2)}(d, j+1)$. Note that $c' \leq C_{r,d}^{(6.4.6)}(c+1)$. We also have $\mathrm{rank}(\mathcal{B}') \geq r(c') \geq c^{(6.1.2)}(d, c'+1)$. Also, by Lemma 6.4.9, for every $b \in \mathbb{F}^{c'}$,

$$\mathbf{Pr}_x[\mathcal{B}'(x) = b] = \frac{1}{|\mathbb{F}|^{c'}} \pm \frac{1}{|\mathbb{F}|^{c'+1}} > 0. \tag{6.3}$$

Since $\mathcal{B}' \succeq_{sem} \mathcal{B}$, there exist $\Gamma : \mathbb{F}^{c'} \to \mathbb{F}$ and for $i \in [c]$, $\Gamma_i : \mathbb{F}^{c'} \to \mathbb{F}$ such that

$$P_i(x) = \Gamma_i(S_1(x), \ldots S_{c'}(x))$$

and

$$Q(x) = \Gamma(S_1(x), \ldots S_{c'}(x)).$$

We next have the following claim.

Claim 6.5.1. *For any $z \in \mathbb{F}^{c'}$. If $\Gamma_i(z) = 0$ for all $i \in [c]$, then $\Gamma(z) = 0$.*

Proof. Suppose $\Gamma_i(z) = 0$ for all $i \in [c]$. Then, by Equation (6.3), there exists $x \in \mathbb{F}^n$ such that $\mathcal{B}'(x) = z$. Thus, for all $i \in [c]$,

$$P_i(x) = \Gamma_i(S_1(x), \ldots S_{c'}(x)) = \Gamma_i(z) = 0.$$

This implies by the hypothesis that

$$\Gamma(z) = Q(x) = \Gamma(S_1(x), \ldots S_{c'}(x)) = 0.$$

□

By Lemma 6.4.17 (and by ensuring $r(j) \geq r_d^{(6.4.17)}(j)$), we have that $\deg(\Gamma_i) \leq d$ for $i \in [c]$ and $\deg(\Gamma) \leq d$. The next step is to obtain $\Lambda_i : \mathbb{F}^{c'} \to \mathbb{F}$, $i \in [c]$ such that,

$$\Gamma(z) = \sum_{i=1}^{c} \Lambda_i(z)\Gamma_i'(z),$$

and $\deg(\Lambda_i) \leq p.O_{d,c}(1)$. This can be done by iterating over $z \in \mathbb{F}^{c'}$ and for each z, set the $\Lambda_i(z)$'s accordingly. Finally compute the Λ_i from its evaluation table. We define $R_i : \mathbb{F}^n \to \mathbb{F}$ for $i \in [c]$, by $R_i(x) = \Lambda_i(S_1(x), \ldots S_{c'}(x))$. Again, it is easy to see that $\deg(R_i) = p.O_{d,c}(1)$. Now, substituting $z = (S_1(x), \ldots S_{c'}(x))$ we get

$$\Gamma(S_1(x), \ldots S_{c'}(x)) = \sum_{i=1}^{c} \Lambda_i(S_1(x), \ldots S_{c'}(x))\Gamma_i'(S_1(x), \ldots S_{c'}(x))$$

which implies

$$Q(x) = \sum_{i=1}^{c} R_i(x) P_i(x).$$

This concludes the proof. □

Computational complexity of the nullstellensatz

We have the following algorithmic version of Theorem 6.1.3.

Corollary 6.5.2 (Algorithmic strong nullstellensatz)**.** *Let $c, d \in \mathbb{N}$. Let $P_1, \ldots P_c \in \mathcal{P}_d(\mathbb{F}^n)$. Let $Q \in \mathcal{P}_d(\mathbb{F}^n)$ such that $Q(x) = 0$ whenever for all i, $P_i(x) = 0$. Then we have an algorithm that performs $n^{p.O_{d,c}(1)}$ field operations that outputs $R_1, \ldots R_c \in \mathcal{P}_D(\mathbb{F}^n)$, $D = O_{d,c}(1)$, such that for r as in Theorem 6.1.3,*

$$Q(x)^r \equiv \sum_{i=1}^{c} R_i(x) P_i(x).$$

Proof. Theorem 6.1.3 guarantees that $R_i \in \mathcal{P}_D(F^n)$. Thus, the terms on the right side have degree $D + d$. Therefore, for each $r \in \mathbb{N}$, we can solve a system of linear equations in $O(c \cdot n^{D+d})$ unknowns by comparing the coefficients on either side and we are guaranteed a solution. Thus, the running time is $(n \cdot \log |\mathbb{F}|)^{p.O_{d,c}(1)}$. □

6.5.2 Ideal Membership Problem

In this subsection, we solve the ideal membership problem in the following setting.

Theorem 6.1.5(Algorithmic Ideal Membership) Let $c, d \in \mathbb{N}$. Let $\beta \in (0, 1)$. Let $P_1, \ldots, P_c, Q \in \mathcal{P}_d(\mathbb{F}^n)$. Let $I = \langle P_1, \ldots, P_c \rangle$. Then we have an algorithm that performs $n^{p.O_{d,c}(1)}$ field operations and decides if $Q \in I$.

Proof. The idea is to directly invoke the nullstellensatz just developed. We set up unknowns for the coefficients of $R_i(x)$ assuming degree $D := D^{(6.1.3)}$. If we obtain a solution to $Q(x) = \sum_i Q_i(x)R_i(x)$ then we output *yes* else output *no*. To prove correctness, we argue as follows. If $Q \in I$, then Q vanishes on the common F-zeroes of the P_i's. Therefore, we can indeed find R_i of degree D by the nullstellensatz. If on the other hand, we we do find such R_i's then trivially $Q \in I$. □

6.5.3 Counting rational points on low degree varieties

We now consider the problem of detecting a rational point in a variety and if so, provide a randomized algorithm that outputs an approximation to the number of rational points in the variety. Along the way, we also show holes in the number of rational points in a variety. The exact problem of detection of rational points is NP hard. See for example [1, 98, 115, 97, 66, 75].

Given polynomials $f_1, \ldots f_c : \mathbb{F}^n \to \mathbb{F}$, let $V_p(f_1, \ldots, f_c) \subseteq \mathbb{F}^n$ denote the set of common zeroes of f_i's, where the subscript p is to emphasize the interest in rational points.

Theorem 6.1.6.[Rational points in varieties] Let $c, d, t, u \in \mathbb{N}$. Let $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ be polynomials of degree d. There is a randomized algorithm that performs $O_{d,c,t,u}(n^d) + |\mathbb{F}|^{O_{d,c,t}(1)}$ field operations and performs the following with probability $1 - \frac{1}{|\mathbb{F}|^t}$.

1. Decide if $V_p(P_1, \ldots, P_c)$ is empty.

2. Output an integer N such that $N = (1 \pm |\mathbb{F}|^{-u})|V_p(P_1, \ldots, P_c)|$.

Proof. Let $\mathcal{B} = \{P_1, \ldots P_c\}$. Applying Lemma 6.4.6, we regularize $\mathcal{B}$ to obtain $\mathcal{B}' = \{S_1, \ldots S_{c'}\}$ with a growth function $r : \mathbb{N} \to \mathbb{N}$ satisfying $r(j) \geq c^{(6.1.2)}(d, j+u)$. Note that $c' \leq C_{r,d}^{(6.4.6)}(c+u)$. We also have $\text{rank}(\mathcal{B}') \geq r(c') \geq c^{(6.1.2)}(d, c'+u)$. Also, by Lemma 6.4.9, for every $b \in \mathbb{F}^{c'}$,

$$\mathbf{Pr}_x[\mathcal{B}'(x) = b] = \frac{1}{|\mathbb{F}|^{c'}} \pm \frac{1}{|\mathbb{F}|^{c'+u}}. \tag{6.4}$$

Since $\mathcal{B}' \succeq_{sem} \mathcal{B}$, there exist $\Gamma_i : \mathbb{F}^{c'} \to \mathbb{F}$ for $i \in [c]$ such that

$$P_i(x) = \Gamma_i(S_1(x), \ldots S_{c'}(x)).$$

Let $S = V_p(\Gamma_1, \ldots, \Gamma_{c'})$ be the set of common zeroes of the Γ_i's. We have the following claim.

Claim 6.5.3. $V_p(\Gamma_1, \ldots, \Gamma_{c'}) = \phi \Leftrightarrow V_p(P_1, \ldots, P_c) = \phi$.

Proof. ($\Rightarrow$ direction) Suppose $V_p(P_1, \ldots, P_c) \neq \phi$. Let $x \in V_p(P_1, \ldots, P_c)$. Then, $z = (S_1(x), \ldots, S_{c'}(x)) \in V_p(\Gamma_1, \ldots, \Gamma_{c'})$.

($\Leftarrow$ direction) Suppose $V_p(\Gamma_1, \ldots, \Gamma_{c'}) \neq \phi$. Let $z \in V_p(\Gamma_1, \ldots, \Gamma_{c'})$. Then, by Equation (6.4), since $\mathbf{Pr}[\mathcal{B}'(x) = z] > 0$, there is an $x \in \mathbb{F}^n$ such that $z = (S_1(x), \ldots, S_{c'}(x))$. Fix an arbitrary $i \in [c]$. Then

$$P_i(x) = \Gamma_i(S_1(x), \ldots, S_{c'}(x)) = \Gamma_i(z) = 0.$$

□

Now, we search over the fixed dimension space in time $|\mathbb{F}|^{c'}$ and this proves the first part of the lemma. We have the following claim which proves the second part of the lemma.

Claim 6.5.4. $|V_p(P_1, \ldots, P_c)| = (1 \pm |\mathbb{F}|^{-u})|\mathbb{F}|^{n-c'}|V_p(\Gamma_1, \ldots, \Gamma_{c'})|$.

Proof. For every $z \in V_p(\Gamma_1, \ldots, \Gamma_{c'})$, by Equation (6.4), the number of points x in $\mathbb{F}^n$ such that $\mathcal{B}(x) = z$ is $(1 \pm |\mathbb{F}|^{-u})|\mathbb{F}|^{n-c'}$. Summing over every such z proves the claim. □

□

Holes in the number of rational points The above lemma states that the number of rational points do not span all possible values. They only lie in the following union of intervals

$$\bigcup_{i=1}^{|\mathbb{F}|^{c'}} \left[i \cdot |\mathbb{F}|^{n-c'} \left(1 - |\mathbb{F}|^{-u}\right), i \cdot |\mathbb{F}|^{n-c'} \left(1 + \mathbb{F}|^{-u}\right) \right].$$

As a special case, we have a strengthening of the Chevellay-Warning theorem in the setting of fixed c, d. The Chevellay-Warning theorem states that if a collection $P_1, \ldots, P_c \in \mathcal{P}_d(\mathbb{F}^n)$ with $dc < n$) has one common solution, it has at least $|\mathbb{F}|$ many solutions.

Corollary 6.5.5. *Let $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ be polynomials of degree d. Then, if the collection has at least one solution, then it has at least $|\mathbb{F}|^{n-O_{c,d}(1)}$ many solutions.*

Such a result for constant size prime fields was proved by the second author in [121]. It is noteworthy to compare the above bound with the Az-Katz theorem [6, 106], which says that the number of solutions is at least $|\mathbb{F}|^{n/d-c}$. More formally,

Theorem 6.5.6 (Ax-Katz theorem)**.** *Let $P_1, \ldots, P_c : \mathbb{F}^n \to \mathbb{F}$ be polynomials of degree d. Then, if the collection has at least one solution, then it has at least $|\mathbb{F}|^{n/d-c}$ solutions.*

6.6 Application: List decoding Reed-Muller codes over large fields

6.6.1 Notation and Preliminaries

Let $\mathbb{F}$ be a prime finite field. A code $\mathcal{C} \subset \{\mathbb{F}^n \to \mathbb{F}\}$ is a subset of functions from $\mathbb{F}^n$ to $\mathbb{F}$, where functions in the code are called codewords. The distance between two functions $f, g : \mathbb{F}^n \to \mathbb{F}$ is the fraction of coordinates where they disagree,

$$\operatorname{dist}(f,g) := \frac{1}{|\mathbb{F}|^n} \left|\{x \in \mathbb{F}^n : f(x) \neq g(x)\}\right|.$$

The minimum distance of a code $\mathcal{C}$ is

$$\operatorname{dist}_{min}(\mathcal{C}) := \min_{f \neq g \in \mathcal{C}} \{\operatorname{dist}(f,g)\}.$$

A code $\mathcal{C}$ is linear if it is a linear subspace over $\mathbb{F}$. For a linear code, $\operatorname{dist}_{min}(\mathcal{C}) = \min_{0 \neq f \in \mathcal{C}} \{\operatorname{dist}(f,0)\}$. For a code $\mathcal{C}$ and a function $g : \mathbb{F}^n \to \mathbb{F}$, the set of codewords at distance at most ρ from g is denoted by

$$B_{\mathcal{C}}(g,\rho) := \{f \in \mathcal{C} : \operatorname{dist}(f,g) \leq \rho\}.$$

The list decoding size of $\mathcal{C}$ at radius ρ is the maximal number of codewords at distance ρ from any possible function,

$$L_{\mathcal{C}}(\rho) := \max_{g:\mathbb{F}^n \to \mathbb{F}} |B_{\mathcal{C}}(g,\rho)|.$$

Recall that $\ell_{\mathbb{F}}(n,d,\rho) := L_{\mathrm{RM}_{\mathbb{F}}(n,d)}(\rho)$. We will use the newly introduced notation for list size for convenience henceforth in this chapter. The Reed-Muller code $\mathrm{RM}_{\mathbb{F}}(n,d)$ is the evaluations of all polynomials $f : \mathbb{F}^n \to \mathbb{F}$ of degree at most d. So using our previous notation, $\mathrm{RM}_{\mathbb{F}}(n,d) = \mathcal{P}_d(\mathbb{F}^n)$. As we assume $d < |\mathbb{F}|$, its minimal distance is given by

$$\mathrm{dist}_{\min}(\mathrm{RM}_{\mathbb{F}}(n,d)) = \min\{\mathbf{Pr}_{x\in\mathbb{F}^n}[f(x)\neq 0] : f : \mathbb{F}^n \to \mathbb{F}, f \neq 0, \deg(f) \leq d\} = 1 - \frac{d}{|\mathbb{F}|}.$$

The main theorem we prove is that Reed-Muller codes, for constant degrees, are list decodable up to their minimal distance. We also extend this to estimate the number of codewords in balls of larger radii.

Theorem 6.1.7. Let $d, s \in \mathbb{N}$. There exists $c = c(d,s)$ such that the following holds. For any prime finite field $\mathbb{F}$ with $|\mathbb{F}| > d$ and any $n \in \mathbb{N}$,

$$L_{\mathrm{RM}_{\mathbb{F}}(n,d)}\left(1 - \frac{d}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right) \leq |\mathbb{F}|^c.$$

Moreover, for any $1 \leq e < d$,

$$L_{\mathrm{RM}_{\mathbb{F}}(n,d)}\left(1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right) \leq |\mathbb{F}|^{c\cdot n^{d-e}}.$$

Both bounds are tight, up to the exact value of $c = c(d,s)$. The proof will follow from a series of propositions which we state next.

Let $\mathrm{RM}_{\mathbb{F}}(n,d,k)$ be a subcode of $\mathrm{RM}_{\mathbb{F}}(n,d)$, which consists of polynomials of degree $\leq d$ and rank $\leq k$. We first reduce the problem of list decoding Reed-Muller codes to list decoding a low rank subcode.

Lemma 6.6.1. *Let $e \leq d, s \in \mathbb{N}$. There is $k = k(d,s)$ such that for any prime field $\mathbb{F}$ with $|\mathbb{F}| > d$ and any $n \in \mathbb{N}$,*

$$L_{\mathrm{RM}_{\mathbb{F}}(n,d)}\left(1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right) \leq |\mathbb{F}|^{2s} \cdot L_{\mathrm{RM}_{\mathbb{F}}(n,d,k)}\left(1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right).$$

So, from now on we restrict our attention to $\mathrm{RM}_{\mathbb{F}}(n,d,k)$. Recall that $\Delta(\mathbb{F})$ is the probability simplex over $\mathbb{F}$, that we naturally embed $\mathbb{F} \subset \Delta(\mathbb{F})$. For $g : \mathbb{F}^n \to \mathbb{F}$ let $p(g) : \mathbb{F}^n \to \Delta(\mathbb{F})$ be this embedding extended to functions. With this notation, for $f, g : \mathbb{F}^n \to \mathbb{F}$ we have $\mathrm{dist}(f,g) = 1 - \langle p(f), p(g) \rangle$. So, from now on we extend our study to functions $\varphi : \mathbb{F}^n \to \Delta(\mathbb{F})$, which can be viewed as randomized functions. The definition of the codewords in $\mathbb{C}$ which are close to a function can be extended to randomized functions following the above discussion:

$$B_{\mathcal{C}}(\varphi, \rho) = \{f \in \mathcal{C} : \langle p(f), \varphi \rangle \geq 1 - \rho\}.$$

Let $\mathcal{F} = \{h_1, \ldots, h_c : \mathbb{F}^n \to \mathbb{F}\}$. We say that φ is $\mathcal{F}$-measureable if $\varphi = \Gamma(\mathcal{F})$ for some function $\Gamma : \mathbb{F}^{|\mathcal{F}|} \to \Delta(\mathbb{F})$. Recall that $\mathbb{E}[\varphi|\mathcal{F}] : \mathbb{F}^n \to \Delta(\mathbb{F})$ as the average of φ with respect to $\mathcal{F}$,

$$\mathbb{E}[\varphi|\mathcal{F}](x) = \mathbb{E}\left[\varphi(y) : y \in \mathbb{F}^n, \mathcal{F}(x) = \mathcal{F}(y)\right].$$

Clearly, $\mathbb{E}[\varphi|\mathcal{F}]$ is $\mathcal{F}$-measurable. Moreover, for any $\xi : \mathbb{F}^n \to \Delta(\mathbb{F})$ which is $\mathcal{F}$-measurable, we have

$$\langle \xi, \varphi \rangle = \langle \xi, E[\varphi|\mathcal{F}] \rangle.$$

We next show that the list decoding problem for low rank codes can be further reduced to the case where the center g is a measurable with respect to a small

polynomial factor of bounded degree. More accurately, it can be list decoded to this latter problem.

Lemma 6.6.2. *Fix $d, k, s \in \mathbb{N}$. There exist $c = c(d, k, s) \in \mathbb{N}$ such that the following holds. Let $\mathbb{F}$ be a prime field with $|\mathbb{F}| > d$ and let $n \in \mathbb{N}$. For any $\varphi : \mathbb{F}^n \to \Delta(\mathbb{F})$ there exists a family of $|\mathbb{F}|^c$ sets of polynomials $\mathcal{F}_i \subset \mathrm{RM}_{\mathbb{F}}(n, d-1)$, $1 \leq i \leq |\mathbb{F}|^c$, of size $|\mathcal{F}_i| \leq c$ each, such that*

$$\forall f \in \mathrm{RM}_{\mathbb{F}}(n, d, k)\ \exists 1 \leq i \leq |\mathbb{F}|^c, |\langle p(f), \varphi \rangle - \langle p(f), \mathbb{E}[\varphi|\mathcal{F}_i] \rangle| \leq |\mathbb{F}|^{-s}.$$

As a corollary, we bound the list decoding size in $\mathrm{RM}_{\mathbb{F}}(n, d, k)$ by the list decoding size when the centers are measurable functions for a system of a few polynomials.

Corollary 6.6.3. *Let $\mathcal{C} = \mathrm{RM}_{\mathbb{F}}(n, d, k)$. Then for any $0 \leq \rho \leq 1$,*

$$B_{\mathcal{C}}(\varphi, \rho) \subset \bigcup_{1 \leq i \leq |\mathbb{F}|^c} B_{\mathcal{C}}(\mathbb{E}[\varphi|\mathcal{F}_i], \rho + |\mathbb{F}|^{-s}).$$

Finally, we prove bounds for the list decoding problem for low rank codes, where the center is measurable with respect to a polynomial factor. In fact, we can even ignore the restriction that the code is low rank, as the restriction on the center is sufficient to obtain the bounds.

Lemma 6.6.4. *Fix $d, s, c \in \mathbb{N}$. There exists $c' = c'(d, s, c)$ such that the following holds. Let $\mathbb{F}$ be a prime field with $|\mathbb{F}| > d$ and let $n \in \mathbb{N}$. Let $\mathcal{F} \subset \mathrm{RM}_{\mathbb{F}}(n, d-1)$ of size $|\mathcal{F}| \leq c$, and let $\varphi : \mathbb{F}^n \to \Delta(\mathbb{F})$ be $\mathcal{F}$-measurable. Then*

$$\left| B_{\mathrm{RM}_{\mathbb{F}}(n,d)} \left(\varphi, 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s} \right) \right| \leq |\mathbb{F}|^{c' \cdot n^{d-e}}.$$

In particular, for any $k \in \mathbb{N}$,

$$\left| B_{\mathrm{RM}_{\mathbb{F}}(n,d,k)} \left(\varphi, 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s} \right) \right| \leq |\mathbb{F}|^{c' \cdot n^{d-e}}.$$

With the above in place, we are ready to prove our main theorem of the section.

Proof of Theorem 6.1.7. Let $\rho := 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}$. By Lemma 6.6.1 there is $k := k(d,s)$ such that

$$L_{\mathrm{RM}_{\mathbb{F}}(n,d)}(\rho) \leq |\mathbb{F}|^{2s} \cdot L_{\mathrm{RM}_{\mathbb{F}}(n,d,k)} \left(\rho + \frac{1}{|\mathbb{F}|^{2s}} \right).$$

Let $g : \mathbb{F}^n \to \mathbb{F}, \varphi = p(g)$. Let $\mathcal{C} = \mathrm{RM}_{\mathbb{F}}(n,d,k)$. Then, by Corollary 6.6.3, for some $c = c(d,s,k)$ we have

$$|B_{\mathcal{C}}(\varphi, \rho)| \leq \sum_{i=1}^{|\mathbb{F}|^c} \left| B_{\mathcal{C}} \left(\mathbb{E}[\varphi|\mathcal{F}_i], 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^{2s}} \right) \right|,$$

where each $\mathcal{F}_i \subset \mathrm{RM}_{\mathbb{F}}(n, d-1)$ of size $|\mathcal{F}_i| \leq c$. Finally, by Lemma 6.6.4, for some $c' = c'(d,s,c)$, we have that for every $1 \leq i \leq |\mathbb{F}|^c$,

$$\left| B_{\mathcal{C}} \left(\mathbb{E}[\varphi|\mathcal{F}_i], 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^{2s}} \right) \right| \leq |\mathbb{F}|^{c' n^{d-e}}.$$

We conclude that

$$L_{\mathrm{RM}_{\mathbb{F}}(n,d)} \left(1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s} \right) \leq |\mathbb{F}|^{2s + c + c' n^{d-e}}.$$

□

We prove Lemma 6.6.1, Lemma 6.6.2 and Lemma 6.6.4 in the following subsections.

6.6.2 Proof of Lemma 6.6.1

We state the Johnson bound first, which provides bounds on the list decoding size for any code, based just on the minimal distance of the code [102].

Lemma 6.6.5 (Johnson bound)**.** *Let* $\mathcal{C} \subseteq \{\mathbb{F}^n \to \mathbb{F}\}$*. Suppose that* $\mathrm{dist}_{min}(\mathcal{C}) \geq 1 - \frac{1}{|\mathbb{F}|} - \varepsilon$*. Then,*

$$L_{\mathcal{C}}\left(1 - \frac{1}{|\mathbb{F}|} - \sqrt{\varepsilon}\right) \leq 1/\varepsilon^2.$$

Proof of Lemma 6.6.1. Set $k = k(d,s) = c^{(6.1.2)}(d, 2s)$. Fix arbitrary $g : \mathbb{F}^n \to \mathbb{F}$. Let

$$L = \left\{ f \in \mathrm{RM}_{\mathbb{F}}(n,d) : \mathrm{dist}(f,g) \leq 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s} \right\}.$$

Let $m = |L|$ and $L = \{f_1, \ldots, f_m\}$. Construct a graph $G = (L, E)$ where $(f_i, f_j) \in E$ if $\mathrm{rank}(f_i - f_j) \leq k$. Let $I \subseteq L$ be a maximal independent set.

Claim 6.6.6. $\mathrm{dist}_{min}(I) \geq 1 - \frac{1}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^{2s}}$.

Proof. Let $f = f_i - f_j \neq 0$ for $f_i, f_j \in I$. Since $\mathrm{rank}(f) > k(d,s) = c^{(6.1.2)}(d, 2s)$, and therefore, $\mathrm{rank}(a \cdot f) > k$ for all $a \in \mathbb{F}, a \neq 0$, we have by Theorem 6.1.2 that $\mathbb{E}\left[e\left(a \cdot f(x)\right)\right] \leq |\mathbb{F}|^{-2s}$. Thus,

$$1 - \mathrm{dist}(f_i, f_j) = \mathbf{Pr}_{x \in \mathbb{F}^n}[f(x) = 0] = \frac{1}{|\mathbb{F}|} \sum_{a \in \mathbb{F}} \mathbb{E}\left[e\left(a \cdot f(x)\right)\right] \leq \frac{1}{|\mathbb{F}|} + \frac{1}{|\mathbb{F}|^{2s}}.$$

□

By the above claim, using the Johnson bound on I, we have that

$$L_I\left(1 - \frac{1}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right) \leq |\mathbb{F}|^{2s}. \tag{6.5}$$

Next, consider any $f \in I$. Say $h_1, \ldots, h_D \in \mathrm{RM}_{\mathbb{F}}(n,d,k)$ are such that $(f+h_i, f) \in E$. As $\mathrm{dist}(g, f+h_i) \leq 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}$, we have that $\mathrm{dist}(g-f, h_i) \leq 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}$. Thus,

$$D \leq L_{RM_{\mathbb{F}}(n,d,k)} \left(1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right). \tag{6.6}$$

Combining Equation (6.5) and Equation (6.6) we conclude that

$$L_{\mathrm{RM}_{\mathbb{F}}(n,d)} \left(1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right) \leq |\mathbb{F}|^{2s} \cdot L_{\mathrm{RM}_{\mathbb{F}}(n,d,k)} \left(1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}\right).$$

□

6.6.3 Proof of Lemma 6.6.2

The proof of Lemma 6.6.2 requires several refinements of $\mathrm{RM}_{\mathbb{F}}(n,d,k)$. First, for $\mathcal{F} \subset \mathrm{RM}_{\mathbb{F}}(n, d-1)$ a family of polynomials of degree $\leq d-1$, define $\mathrm{RM}_{\mathbb{F}}(n,d,k,\mathcal{F})$ to be the family of degree d polynomials, which can be decomposed as a function of the polynomials in $\mathcal{F}$, and k additional polynomials of degree $\leq d-1$.

For $\mathbf{k} = (k_1, \ldots, k_{d-1}) \in \mathbb{N}^{d-1}$ let $|\mathbf{k}| = \sum k_i$. The code $\mathrm{RM}_{\mathbb{F}}(n,d,\mathbf{k},\mathcal{F})$ is a subcode of $\mathrm{RM}_{\mathbb{F}}(n,d,|\mathbf{k}|,\mathcal{F})$, defined as family of degree d polynomials, which can be decomposed as a function of the polynomials in $\mathcal{F}$, and $|\mathbf{k}|$ additional polynomials, with k_i polynomials of degree i, for $1 \leq i \leq d-1$. The following statement of the theorem allows for a streamlined inductive proof.

Theorem 6.6.7. *Fix $d, s \in \mathbb{N}, \mathbf{k} \in \mathbb{N}^{d-1}, \mathcal{F} \subset \mathrm{RM}_{\mathbb{F}}(n, d-1)$ and let $\mathcal{C} = \mathrm{RM}_{\mathbb{F}}(n,d,\mathbf{k},\mathcal{F})$. There exist $c = c(d,\mathbf{k},s,|\mathcal{F}|) \in \mathbb{N}$ such that the following holds. For any $\varphi : \mathbb{F}^n \to \Delta(\mathbb{F})$ there exists a family of $|\mathbb{F}|^c$ sets of polynomials $\mathcal{F}_i \subset \mathrm{RM}_{\mathbb{F}}(n, d-1)$,*

$1 \leq i \leq |\mathbb{F}|^c$, *of size* $|\mathcal{F}_i| \leq c$ *each, such that*

$$\forall f \in \mathcal{C}\ \exists 1 \leq i \leq |\mathbb{F}|^c, |\langle p(f), \varphi\rangle - \langle p(f), \mathbb{E}[\varphi|\mathcal{F}_i]\rangle| \leq |\mathbb{F}|^{-s}.$$

The simplex. Recall that for $f : \mathbb{F}^n \to \mathbb{F}$ we have $p(f) : \mathbb{F}^n \to \Delta(\mathbb{F})$. Define $q(f) := p(f) - \frac{1}{|\mathbb{F}|}$, so that $\sum_{y\in\mathbb{F}} q(f)(x)_y = 0$ for all $x \in \mathbb{F}^n$. For $a \in \mathbb{F}^n, b \in \mathbb{F}$, define $\ell_{a,b} : \mathbb{F}^n \to \mathbb{F}$ by $\ell_{a,b}(x) = \langle a, x\rangle + b$. We prove the following analogue of Fourier expansion over the simplex. We will refer to it as the Fourier simplex decomposition.

Lemma 6.6.8. *Let* $g : \mathbb{F}^n \to \mathbb{F}$*. Then,*

$$q(g)(x) = \sum_{a\in\mathbb{F}^n, 0\neq b\in\mathbb{F}} \alpha_{a,b} q(\ell_{a,b})(x),$$

where

$$\alpha_{a,b} = \langle q(g), q(\ell_{a,b})\rangle - \langle q(g), q(\ell_{a,0})\rangle$$

are unique and satisfy $\alpha_{a,b} \in [-1, 1]$*. We denote the* $\alpha_{a,b}$ *by* $\widehat{q(g)}(a, b)$*.*

Proof. We first construct a basis for the subspace $V \subseteq (\mathbb{R}^{|\mathbb{F}|})^{|\mathbb{F}|^n}$ defined as follows. We index the coordinates of $v \in (\mathbb{R}^{|\mathbb{F}|})^{|\mathbb{F}|^n}$ as $v_{x,y}$, with $x \in \mathbb{F}^n, y \in \mathbb{F}$. Then

$$V = \left\{ v_{x,y} \in \mathbb{R}^{|\mathbb{F}|^{n+1}} : \sum_{y\in\mathbb{F}} v_{x,y} = 0\ \forall x \in \mathbb{F}^n \right\}.$$

Note that $q(g) \in V$ for any $g : \mathbb{F}^n \to \mathbb{F}$. Also, $\dim(V) = |\mathbb{F}|^n(|\mathbb{F}| - 1)$. We next establish that the set of vectors

$$I = \{q(\ell_{a,b}) : a \in \mathbb{F}^n, 0 \neq b \in \mathbb{F}\} \subseteq V$$

is a basis for V. First, note that $|I| = |\mathbb{F}|^n(|\mathbb{F}| - 1)$. To prove linear independence of I, suppose that

$$\Lambda := \sum_{a \in \mathbb{F}^n, 0 \neq b \in \mathbb{F}} \alpha_{a,b} q(\ell_{a,b}) = 0. \tag{6.7}$$

Let $V_a = \text{span}\{q(\ell_{a,b}) : 0 \neq b \in \mathbb{F}\}$. Then, by Equation (6.7), $\sum_a v_a = 0$ where $v_a = \sum_{0 \neq b \in \mathbb{F}} \alpha_{a,b} q(\ell_{a,b}) \in V_a$. We now note that $\langle v_a, v_{a'} \rangle = 0$ if $a \neq a'$. Indeed,

$$\langle v_a, v_{a'} \rangle = \left\langle \sum_{b \neq 0} \alpha_{a,b} q(\ell_{a,b}), \sum_{b' \neq 0} \alpha_{a',b'} q(\ell_{a',b'}) \right\rangle = 0,$$

since for any $a \neq a' \in \mathbb{F}^n$ and any $b, b' \in \mathbb{F}$,

$$\langle q(\ell_{a,b}), q(\ell_{a',b'}) \rangle = \mathbf{Pr}_{x \in \mathbb{F}^n} \left[\langle a, x \rangle + b = \langle a', x \rangle + b' \right] - \frac{1}{|\mathbb{F}|} = 0.$$

In particular, $\langle v_a, v_a \rangle = 0$ for all $a \in \mathbb{F}^n$ which implies that $v_a = 0$. Fix an arbitrary $a \in \mathbb{F}^n$. We now show that $v_a = \sum_{b \neq 0} \alpha_{a,b} q(\ell_{a,b}) = 0$ implies that $\alpha_{a,b} = 0$ for all $b \neq 0$. Indeed, fix $x \in \mathbb{F}^n$ such that $\langle a, x \rangle = 0$. Then $v_a(x)_y = \alpha_{a,y} - \sum_{b \neq 0} \alpha_{a,b}$ if $y \neq 0$, and $v_a(x)_0 = -\sum_{b \neq 0} \alpha_{a,b}$. As we have that $v_a(x)_y = 0$ for all $y \in \mathbb{F}$, it must be that $\alpha_{a,b} = 0$ for all $0 \neq b \in \mathbb{F}$.

Thus, I indeed forms a basis for V. The uniqueness of the $\alpha_{a,b}$ follows from the linear independence of I. So far, we have established that

$$q(g)(x) = \sum_{a \in \mathbb{F}^n, 0 \neq b \in \mathbb{F}} \alpha_{a,b} q(\ell_{a,b})(x). \tag{6.8}$$

Using the simple fact that

$$\langle q(\ell_{a,b}), q(\ell_{a',b'}) \rangle = \mathbf{Pr}[\langle a, x \rangle + b = \langle a', x \rangle + b'] - \frac{1}{|\mathbb{F}|},$$

we record the following observation.

$$\langle q(\ell_{a,b}), q(\ell_{a',b'})\rangle = \begin{cases} 0 & \text{if } a \neq a' \\ 1 - \frac{1}{|\mathbb{F}|} & \text{if } a = a', b = b' \\ -\frac{1}{|\mathbb{F}|} & \text{if } a = a', b \neq b' \end{cases}$$

Taking inner product on both sides of Equation (6.8) with $q(\ell_{a,b})$ we get,

$$\langle q(g), q(\ell_{a,b})\rangle = \left(1 - \frac{1}{|\mathbb{F}|}\right)\alpha_{a,b} - \frac{1}{|\mathbb{F}|}\left(\sum_{b' \neq 0, b} \alpha_{a,b'}\right) = \alpha_{a,b} - \frac{1}{|\mathbb{F}|}\sum_{b' \neq 0}\alpha_{a,b'}.$$

Summing for all $b \neq 0$, we obtain that

$$\sum_{b \neq 0}\langle q(g), q(\ell_{a,b})\rangle = \frac{1}{|\mathbb{F}|}\sum_{b' \neq 0}\alpha_{a,b'}. \tag{6.9}$$

Thus,

$$\alpha_{a,b} = \langle q(g), q(\ell_{a,b})\rangle + \sum_{b' \neq 0}\langle q(g), q(\ell_{a,b'})\rangle. \tag{6.10}$$

Next, we observe that $\sum_{b \in \mathbb{F}} q(\ell_{a,b}) = 0$. This is since

$$\sum_{b \in \mathbb{F}} q(\ell_{a,b})_{x,y} = \sum_{b \in \mathbb{F}}\left(\mathbf{Pr}[\langle a, x\rangle + b = y] - \frac{1}{|\mathbb{F}|}\right) = 1 - 1 = 0.$$

So we have

$$\alpha_{a,b} = \langle q(g), q(\ell_{a,b})\rangle - \langle q(g), q(\ell_{a,0})\rangle.$$

Since $\langle q(g), q(\ell_{a,b})\rangle \in [-\frac{1}{|\mathbb{F}|}, 1 - \frac{1}{|\mathbb{F}|}]$ for all $b \in \mathbb{F}$, we obtain that $\alpha_{a,b} \in [-1, 1]$. This finishes the proof □

Weak regularity on the simplex. We prove the following lemma. In the following, X, Y are arbitrary finite sets, where we will later apply the lemma to $X = \mathbb{F}^n, Y = \mathbb{F}$. The proof is similar to Frieze-Kannan weak regularity [62] but generalized to the simplex.

Lemma 6.6.9. *Let $\varepsilon > 0$ be arbitrary. Let $\varphi : X \to \Delta(Y)$ be arbitrary. Let $\mathcal{F}$ be a collection of functions $f : X \to Y$. Then, there exist $f_1, \ldots, f_c \in \mathcal{F}$, $c \leq 1/\varepsilon^2$ such that*

$$\varphi = \frac{1}{|Y|} + \sum_{i=1}^{c} \alpha_i q(f_i) + h,$$

where $|\alpha_i| \leq 1$ and h satisfies that for all $f \in \mathcal{F}$,

$$|\langle h, q(f) \rangle| \leq \varepsilon.$$

Proof. Let $\varphi' = \varphi - \frac{1}{|Y|}$. We will define a sequence of functions $\varphi_i \in \mathbb{F}^n \to \mathbb{R}^{\mathbb{F}}$. Initialize $\varphi_0 := 0$. Given φ_i, if there exists $f_i \in \mathcal{F}$ such that $\langle \varphi' - \varphi_i, q(f_i) \rangle = \alpha_i$ where $|\alpha_i| > \varepsilon$, set $\varphi_{i+1} := \varphi_i + \alpha_i q(f_i)$. We show that the process terminates quickly. To that end, define $\delta_i := ||\varphi' - \varphi_i||_2^2$. Then

$$\begin{aligned}
\delta_{i+1} &= ||\varphi' - \varphi_i - \alpha_i q(f_i)||_2^2 \\
&= \delta_i + \alpha_i^2 ||q(f_i)||_2^2 - 2\langle \varphi' - \varphi_i, \alpha_i q(f_i) \rangle \\
&= \delta_i + \alpha_i^2 (1 - 1/|Y|) - 2\alpha_i^2 \\
&\leq \delta_i - \alpha_i^2 \\
&\leq \delta_i - \varepsilon^2.
\end{aligned}$$

Additionally, $\delta_0 = ||\varphi'||_2^2 \leq 1$ and $\delta_i \geq 0$ for all i. Thus, the process terminates after $\leq 1/\varepsilon^2$ steps. At the end of the process, we have

$$\varphi' = \sum_{i=1}^{c} \alpha_i q(f_i) + h,$$

where h satisfies that for all $f \in \mathcal{F}$, $|\langle h, q(f) \rangle| \leq \varepsilon$ and $|\alpha_i| \leq \sqrt{\delta_i - \delta_{i+1}} \leq 1$. □

Proof of Theorem 6.6.7. The proof is by induction on $d, s, \mathbf{k}, |\mathcal{F}|$. For $\mathbf{k}$, we use the lexicographic order on $\mathbb{N}^{d-1}$ which is well founded to define a Noetherian induction. Let $\mathcal{C} = \mathrm{RM}_{\mathbb{F}}(n, d, \mathbf{k}, \mathcal{F})$, and fix $\varphi : \mathbb{F}^n \to \Delta(\mathbb{F})$ and $s \geq 1$, and set $e = |\mathcal{F}|$.

We first argue that we may assume that $\mathcal{F}$ is regular. For a rank function $R_1 : \mathbb{N} \to \mathbb{N}$ to be determined later (as a function of $d, \mathbf{k}, s$), regularize $\mathcal{F}$ to obtain an R_1-regular factor $\mathcal{F}'$. Note that

$$\mathrm{RM}_{\mathbb{F}}(n, d, \mathbf{k}, \mathcal{F}) \subset \mathrm{RM}_{\mathbb{F}}(n, d, \mathbf{k}, \mathcal{F}').$$

Thus, we may instead study $\mathrm{RM}_{\mathbb{F}}(n, d, \mathbf{k}, \mathcal{F}')$. So, we simply assume from now on that $\mathcal{F}$ is r_1-regular for some $r_1 = R_1(d, \mathbf{k}, s, |\mathcal{F}|)$ to be determined later.

Let $f \in \mathcal{C}$. By definition, we can decompose f as

$$f = \Gamma(\mathcal{H}, \mathcal{F}),$$

where $\mathcal{H} = \{h_1, \ldots, h_k\}$ is a family of k polynomials of degree $\leq d - 1$ and where $\Gamma : \mathbb{F}^{k+e} \to \mathbb{F}$ is some function. We argue that we can also assume that $\mathcal{H} \cup \mathcal{F}$ is regular. If $\mathcal{H} \cup \mathcal{F}$ is not $r_2 = r_1 - 1$ regular, then

$$\mathrm{rank}\left(\sum a_i h_i + \sum b_i f_i\right) \leq r_2,$$

for some $a_i, b_i \in \mathbb{F}$, not all zero. Let d' be the maximal degree of a polynomial appearing in the linear combination with a nonzero coefficient. It cannot be that all these polynomials are in $\{f_i\}$, as we assumed that the rank of $\mathcal{F}$ is at least r_1. So, $a_i \neq 0$ for some i where $\deg(h_i) = d'$. This means h_i can be expressed as a function of the other polynomials in $\mathcal{H} \cup \mathcal{F}$, and an additional set $\mathcal{H}'$ of $r_1 - 2$ polynomials of

degrees $\leq d'-1$. So, if we construct $\mathbf{k}'$ from $\mathbf{k}$ by reducing the number of polynomials of degree d' by one, and increasing the number of polynomials of degrees $\leq d'-1$ by r_2, then in fact we have

$$f \in \mathrm{RM}_{\mathbb{F}}(n, d, \mathbf{k}', \mathcal{F}).$$

Thus, we may apply the theorem by induction in order to handle these polynomials, since $\mathbf{k}' < \mathbf{k}$ in the lexicographic order. So, we assume from now on that $\mathcal{F} \cup \mathcal{H}$ is r_2-regular.

Let $\psi = \mathbb{E}[\varphi|\mathcal{F}]$. We will include $\mathcal{F}$ as one of our sets $\mathcal{F}_i$, and hence handle any f for which $|\langle q(f), \psi\rangle - \langle q(f), \varphi\rangle| \leq |\mathbb{F}|^{-s}$. So, from now on we consider only f for which $|\langle q(f), \psi\rangle - \langle q(f), \varphi\rangle| \geq |\mathbb{F}|^{-s}$. Decomposing Γ to its Fourier simplex decomposition (Lemma 6.6.8), and applying this to decompose f, we obtain that

$$q(f)(x) = \sum_{a \in \mathbb{F}^k, b \in \mathbb{F}^e, 0 \neq c \in \mathbb{F}} \widehat{\Gamma}(a,b,c) \cdot q\left(\sum a_i h_i(x) + \sum b_i f_i(x) + c\right),$$

where $|\widehat{\Gamma}(a,b,c)| \leq 1$. Note that whenever $a = 0$, we have

$$\left\langle q\left(\sum b_i f_i + c\right), \varphi\right\rangle = \left\langle q\left(\sum b_i f_i + c\right), \psi\right\rangle,$$

since $\sum b_i f_i + c$ is $\mathcal{F}$-measurable. Hence, there must exist $0 \neq a \in \mathbb{F}^k, b \in \mathbb{F}^e, c \neq 0$, such that

$$\left|\left\langle q\left(\sum a_i h_i(x) + \sum b_i f_i(x) + c\right), \varphi - \psi\right\rangle\right| \geq |\mathbb{F}|^{-(s+k+e+1)}.$$

As $\mathbb{E}[\varphi - \psi] = 0$ we equivalently have

$$\left|\left\langle p\left(\sum a_i h_i(x) + \sum b_i f_i(x) + c\right), \varphi - \psi\right\rangle\right| \geq |\mathbb{F}|^{-(s+k+e+1)}.$$

Next, we decompose by Lemma 6.6.9 both φ and ψ, and subtract the decompositions obtain that

$$\varphi - \psi = \sum_{t=1}^{\ell} \gamma_t \cdot q(w_t) + \xi,$$

where $\gamma_t \in [-1, 1]$, $w_t \in \mathcal{C}$, $\xi : \mathbb{F}^n \to \mathbb{R}^{\mathbb{F}}$ satisfies that $|\langle \xi, f \rangle| \leq |\mathbb{F}|^{-2(s+k+e+1)}$ for all $f \in \mathcal{C}$, and $\ell \leq |\mathbb{F}|^{4(s+k+e+1)}$. There must exist $t \in [\ell]$ such that

$$\left| \left\langle p\left(\sum a_i h_i(x) + \sum b_i f_i(x) + c \right), q(w_t) \right\rangle \right| \geq |\mathbb{F}|^{-5(s+k+e)}.$$

As $w_t \in \mathcal{C}$ we can decompose it as as a function of k polynomials of degree $\leq d-1$ and $\mathcal{F}$. Let $R_3 : \mathbb{N} \to \mathbb{N}$ be large enough to be determined later (as a function of $d, \mathbf{k}, s$). We regularize these polynomials to be R_3-regular, and obtain a collection of $\leq c_1(d, \mathbf{k}, |\mathcal{F}|, s)$ polynomials. We choose R_1 large enough so that $R_1(e) > c_1(d, \mathbf{k}, e, s)$. This ensures that $\mathcal{F}$ does not change in the regularization process. Hence we have

$$w_t = \Gamma_t(\mathcal{H}_t \cup \mathcal{F}),$$

where $\mathcal{H}_t \cup \mathcal{F}$ is R_3-regular, $|\mathcal{H}_t| = k_t \leq c_1(d, \mathbf{k}, e, s)$, $\mathcal{H}_t = \{h_{t,1}, \ldots, h_{t,k_t}\}$ and $\Gamma_t : \mathbb{F}^{k_t+e} \to \mathbb{F}$ is some function. Decomposing Γ_t to its Fourier decomposition, and applying this to decompose w_t, we obtain that

$$q(w_t)(x) = \sum_{a \in \mathbb{F}^{k_t}, b \in \mathbb{F}^e, c \neq 0} \widehat{\Gamma}_t(a, b, c) \cdot q\left(\sum a_i h_{t,i}(x) + \sum b_i f_i(x) + c \right).$$

So, there must exist $a' \in \mathbb{F}^{k_t}, b' \in \mathbb{F}^e, c' \neq 0$ such that

$$\left| \left\langle p\left(\sum a_i h_i(x) + \sum b_i f_i(x) + c \right), q\left(\sum a'_i h_{t,i}(x) + \sum b'_i f_i(x) + c' \right) \right\rangle \right| \geq |\mathbb{F}|^{-5(s+k+e+1)-(k_t+e+1)},$$

which equivalently means that, for $b_i'' = b_i - b_i'$ and $c'' = c - c'$, that

$$\left| \mathbf{Pr}_{x \in \mathbb{F}^n} \left[\sum a_i h_i(x) - \sum a_i' h_{t,i}(x) + \sum b_i'' f_i(x) + c'' = 0 \right] - \frac{1}{|\mathbb{F}|} \right| \geq |\mathbb{F}|^{-(5(s+k+e+1)+(k_t+e+1))}.$$

This implies that

$$\operatorname{rank}\left(\sum a_i h_i(x) - \sum a_i' h_{t,i}(x) + \sum b_i'' f_i(x) \right) \leq r_4 = r_4(d,k,e,s).$$

Let d' be the maximal degree of a polynomial appearing in the linear combination with a nonzero coefficient. By choosing R_3 large enough, we guarantee that it cannot be the case that all the polynomials of degree d' are in $\mathcal{H}_t \cup \mathcal{F}$. So, $a_i \neq 0$ for some i such that $\deg(h_i) = d'$. So, we can express h_i as a function of all the other polynomials in $\mathcal{H} \cup \mathcal{H}_t \cup \mathcal{F}$, and an additional set of r_4 polynomials of degree $\leq d' - 1$. Thus, we define $\mathcal{F}_t = \mathcal{F} \cup \mathcal{H}_t$, and construct $\mathbf{k}'$ from $\mathbf{k}$ by decreasing the number of polynomials of degree d' by one, and increase the number of polynomials of any lower degree by r_4, then $\mathbf{k}' < \mathbf{k}$ and we obtain that in fact

$$f \in \mathrm{RM}_{\mathbb{F}}(n, d, \mathbf{k}', \mathcal{F}_t).$$

Crucially, the sets $\mathcal{F}_t$ were obtained depending only on φ and $\mathcal{F}$. Thus, we can apply the theorem by induction to each of them. Let $\{\mathcal{F}_{t,i} : 1 \leq i \in |\mathbb{F}|^{c'}\}$ be the sets guaranteed by the theorem, where $c' \leq c(d, \mathbf{k}', s, \mathcal{F}_t)$. We conclude the proof by taking their union, which has size $\leq |\mathbb{F}|^{4(s+k+e)} \cdot |\mathbb{F}|^{c(d,\mathbf{k}',s,|\mathcal{F}_t|)}$, which is bounded by $|\mathbb{F}|^c$ for a large enough $c = c(d, \mathbf{k}, s, |\mathcal{F}|)$. □

6.6.4 Proof of Lemma 6.6.4

The proof of Lemma 6.6.4 is similar to the authors' previous work [35]. Let $\mathcal{F} = \{h_1, \ldots, h_c\}$ a family of polynomials of degree $\leq d - 1$, and let $\varphi : \mathbb{F}^n \to \Delta(\mathbb{F})$

be $\mathcal{F}$-measurable. Let $\rho := 1 - \frac{e}{|\mathbb{F}|} - \frac{1}{|\mathbb{F}|^s}$. Fix $f \in \mathcal{C}$ such that

$$\langle p(f), \varphi \rangle \geq 1 - \rho.$$

For $a \in \mathbb{F}^c$ define

$$A_a := \{x \in \mathbb{F}^n : h_1(x) = a_1, \ldots, h_c(x) = a_c\}.$$

Define $\Gamma_f : \mathbb{F}^c \to \mathbb{F}$ by setting $\Gamma_f(a)$ to be the most common value f attains on A_a. Then

$$\begin{aligned}
\mathbf{Pr}[f(x) = \Gamma_f(h_1(x), \ldots, h_c(x))] &= \sum_{a \in \mathbb{F}^c} \mathbf{Pr}[x \in A_a] \cdot \max_{y^* \in \mathbb{F}} \mathbf{Pr}[f(x) = y^* | x \in A_a] \\
&\geq \sum_{a \in \mathbb{F}^c} \mathbf{Pr}[x \in A_a] \cdot \mathbb{E}\left[\langle p(f), \varphi \rangle | x \in A_a\right] \\
&= \mathbb{E}[[\langle p(f), \varphi \rangle] \\
&\geq 1 - \rho.
\end{aligned}$$

Let $r_1, r_2 : \mathbb{N} \to \mathbb{N}$ be two non decreasing functions to be specified later, and let $C_{r,d}^{(6.4.6)}$ be as given in Lemma 6.4.6. We will require that for all $m \geq 1$,

$$r_1(m) \geq r_2(C_{r_2,d}^{(6.4.6)}(m+1)) + C_{r_2,d}^{(6.4.6)}(m+1) + 1. \tag{6.11}$$

Let $\mathcal{B}$ be the factor defined by $\mathcal{F}$. As a first step, we r_1-regularize $\mathcal{F}$ by Lemma 6.4.6. This gives an r_1-regular factor $\mathcal{B}'$ of degree at most d, defined by polynomials $\mathcal{F}' = \{h_1', \ldots, h_{c'}' : \mathbb{F}^n \to \mathbb{F}\}$, such that $\mathcal{B}' \succeq_{sem} \mathcal{B}$, $c' \leq C_{r_1,d}^{(6.4.6)}(c)$ and $\mathrm{rank}(\mathcal{B}') \geq r_1(c')$. Let $G_f : \mathbb{F}^{c'} \to \mathbb{F}$ be defined such that

$$G_f(h_1'(x), \ldots, h_{c'}'(x)) = \Gamma_f(h_1(x), \ldots, h_c(x)).$$

Then

$$\mathbf{Pr}[G_f(h'_1(x), h'_2(x), \ldots, h'_{c'}(x)) = f(x)] \geq 1 - \rho. \tag{6.12}$$

Appealing again to Lemma 6.4.6, we r_2-regularize $\mathcal{B}_f := \mathcal{B}' \cup \{f\}$. We get an r_2-regular factor $\mathcal{B}'' \succeq_{syn} \mathcal{B}'$ defined by the collection $\mathcal{F}'' = \{h'_1, \ldots, h'_{c'}, h''_1, \ldots, h''_{c''}\} \subseteq \mathrm{RM}_{\mathbb{F}}(n, d-1)$. Note that it is a syntactic refinement of $\mathcal{B}'$ as by our choice of r_1,

$$\mathrm{rank}(\mathcal{B}') \geq r_1(c') \geq r_2(C^{(6.4.6)}_{r_2,d}(c'+1)) + C^{(6.4.6)}_{r_2,d}(c'+1) + 1 \geq r_2(|\mathcal{B}''|) + |\mathcal{B}''| + 1.$$

We will choose r_2 such that for all $m \geq 1$,

$$r_2(m) = \max\left(r^{(6.4.9)}(d, 2s+m), r^{(6.4.16)}(m, d, d)\right). \tag{6.13}$$

Since f is measurable with respect to $\mathcal{B}''$, there exists $F : \mathbb{F}^{c'+c''} \to \mathbb{F}$ such that

$$f(x) = F(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)).$$

As will see soon, our goal is to analyze the structure of F. We next show that we can have each polynomial in the factor have a disjoint set of inputs. Let $r \in \mathbb{N}$ be large enough to be determined later. Let $n_1 = r\sum_{i=1}^{c'} \deg(h'_i)$ and $n_2 = r\sum_{i=1}^{c''} \deg(h''_i)$. Define $y \in \mathbb{F}^{n_1}$ indexed as $y_{i,j,k}$, with $i \in [c'], j \in [r], k \in [\deg(h'_i)]$, and define $z \in \mathbb{F}^{n_2}$ indexed as $z_{i,j,k}$, with $i \in [c''], j \in [r], k \in [\deg(h''_i)]$. Define new polynomials $\widetilde{h'_i}(y), \widetilde{h''_i}(z)$ as follows:

$$\widetilde{h'_i}(y) = \sum_{j=1}^{r} \prod_{k=1}^{\deg(h'_i)} y_{i,j,k} \qquad \forall i \in [c'],$$

$$\widetilde{h''_i}(z) = \sum_{j=1}^{r} \prod_{k=1}^{\deg(h''_i)} z_{i,j,k} \qquad \forall i \in [c''].$$

Note that the polynomials $\{h'_i : i \in [c']\}, \{h''_i : i \in [c'']\}$ are defined over disjoint sets of variables, and that $\deg(\widetilde{h'_i}) = \deg(h'_i)$ and $\deg(\widetilde{h''_i}) = \deg(h''_i)$. Define new functions $\widetilde{f} : \mathbb{F}^{n_1+n_2} \to \mathbb{F}$ and $\widetilde{g} : \mathbb{F}^{n_1} \to \mathbb{F}$ as follows:

$$\widetilde{f}(y,z) = F(\widetilde{h'_1}(y), \ldots, \widetilde{h'_{c'}}(y), \widetilde{h''_1}(z), \ldots, \widetilde{h''_{c''}}(z)),$$
$$\widetilde{g}(y) = G_f(\widetilde{h'_1}(y), \ldots, \widetilde{h'_{c'}}(y)).$$

Claim 6.6.10. *For a large enough $r = r(d, c', c'', s)$ it holds that $\deg(\widetilde{f}) \leq d$ and*

$$\left|\mathbf{Pr}_{y\in\mathbb{F}^{n_1}, z\in\mathbb{F}^{n_2}}[\widetilde{f}(y,z) = \widetilde{g}(y)] - \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))]\right| \leq \frac{1}{|\mathbb{F}|^{s+1}}.$$

Proof. The bound $\deg(\widetilde{f}) \leq \deg(f) \leq d$ follows from Lemma 6.4.16 since $r_2(|\mathcal{F}''|) \geq r_d^{(6.4.16)}(|\mathcal{F}''|)$. To establish the bound on $\mathbf{Pr}[\widetilde{f} = \widetilde{g}]$, for each $a \in \mathbb{F}^{c'+c''}$ let

$$p_1(a) = \mathbf{Pr}_{x\in\mathbb{F}^n}[(h'_1(x), \ldots, h'_{c'}(x), h''_1(x), \ldots, h''_{c''}(x)) = a].$$

Applying Lemma 6.4.9 and since our choice of r_2 satisfies $\mathrm{rank}(\mathcal{F}'') \geq r^{(6.4.9)}(d, s + 2|\mathcal{F}''|)$, we have that p_1 is nearly uniform over $\mathbb{F}^{c'+c''}$,

$$p_1(a) = \frac{1 \pm |\mathbb{F}|^{-2s}}{|\mathbb{F}|^{c'+c''}}.$$

Similarly, let

$$p_2(a) = \mathbf{Pr}_{y\in\mathbb{F}^{n_1}, z\in\mathbb{F}^{n_2}}[(\widetilde{h'_1}(y), \ldots, \widetilde{h'_{c'}}(y), \widetilde{h''_1}(z), \ldots, \widetilde{h''_{c''}}(z)) = a].$$

For r large enough, as the polynomials are evaluated on disjoint variables, it also holds that

$$p_2(a) = \frac{1 \pm |\mathbb{F}|^{-2s}}{|\mathbb{F}|^{c'+c''}}.$$

For $a \in \mathbb{F}^{c'+c''}$, let $a' \in \mathbb{F}^{c'}$ be the restriction of a to first c' coordinates, $a' = (a_1, \ldots, a_{c'})$. Thus

$$\begin{aligned}\mathbf{Pr}_{y\in\mathbb{F}^{n_1}, z\in\mathbb{F}^{n_2}}[\widetilde{f}(y,z) = \widetilde{g}(y)] &= \sum_{a\in\mathbb{F}^{c'+c''}} p_2(a) 1_{F(a)=G_f(a')} \\ &= \sum_{a\in\mathbb{F}^{c'+c''}} p_1(a) 1_{F(a)=G_f(a')} \pm |\mathbb{F}|^{-2s} \\ &= \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x), h'_2(x), \ldots, h'_c(x))] \pm |\mathbb{F}|^{-2s}.\end{aligned}$$

□

So, we obtain that

$$\mathbf{Pr}_{y\in\mathbb{F}^{n_1}, z\in\mathbb{F}^{n_2}}[\widetilde{f}(y,z) = \widetilde{g}(y)] \geq \mathbf{Pr}_{x\in\mathbb{F}^n}[f(x) = G_f(h'_1(x), \ldots, h'_{c'}(x))] - |\mathbb{F}|^{-2s} \geq \frac{e}{|\mathbb{F}|} + |\mathbb{F}|^{-2s}.$$

In the remaining part of the proof, we show that $\deg(\widetilde{h''_j}) \leq d - e$. Since,

$$\operatorname{rank}(\mathcal{B}'') \geq r_2(|\mathcal{B}''|) \geq r^{(6.4.16)}(|\mathcal{B}''|, d, d),$$

this implies that $\deg(F) \leq d$ by Lemma 6.4.17. This immediately proves that the number of $f \in B_{\mathcal{C}}(\varphi, \rho)$ is bounded by

$$|B_{\mathcal{C}}(\varphi,\rho)| \leq (\#\text{ of } F)(\#\text{ of } h''_1, \ldots, h''_{c''}) \leq |\mathbb{F}|^{(c'+c'')^d} |\mathbb{F}|^{O\left(c'' n^{d-e}\right)} = |\mathbb{F}|^{O_{d,s,c}\left(n^{d-e}\right)}.$$

To conclude, we prove the following.

Claim 6.6.11. *$\deg(\widetilde{h''_i}) \leq d - e$ for all $i \in [c'']$.*

Proof. To simplify notations, let $h = h''_i$, $n' = r \cdot \deg(h''_i)$, let $w' \in \mathbb{F}^{n'}$ denote the inputs to $\widetilde{h''_i}$, namely $\{z_{i,j,k} : j \in [r], k \in \deg(h''_i)\}$, and let $w'' \in \mathbb{F}^{n_1+n_2-n'}$ denote all

the remaining inputs from y, z. Let $n'' = n_1 + n_2 - n'$. Then we have

$$\widetilde{f}(w', w'') = \Gamma'(w'', h(w')), \quad \widetilde{g}(w'') = \Gamma''(w'').$$

Let $d_0 := \deg(h)$, where our goal is to prove that $d_0 \leq d - e$. Note that as $\deg(\widetilde{f}) \leq d$ by Claim 6.6.10, we must have $\deg(\Gamma') \leq d$. Thus we can expand

$$\Gamma'(w''_1, \ldots, w''_{n''}, t) = \sum_{i=0}^{d'} q_i(w''_1, \ldots, w''_{n''}) t^i,$$

where $d' \leq d$ and $q_{d'} \neq 0$. Moreover, by choosing $r > d^2$, we have that $\deg(h^i) = i \cdot \deg(h)$ for any $i \leq d$. Thus, we have $\deg(q_i) \leq d - i \cdot d_0$. We have

$$\mathbf{Pr}[\widetilde{f}(w'', h(w')) = \widetilde{g}(w'')] = \mathbf{Pr}\left[(q_0 - \Gamma'')(w'') + \sum_{i=1}^{d'} q_i(w'') h(w)^i = 0\right].$$

We upper bound this probability as a combination of two terms. Consider any fixing of w''. The probability that $q_{d'}(w'') = 0$ is bounded by

$$\mathbf{Pr}[q_{d'}(w'') = 0] \leq \frac{\deg(q_{d'})}{|\mathbb{F}|} \leq \frac{d - d'd_0}{|\mathbb{F}|}.$$

Otherwise, we have $q_{d'}(w'') \neq 0$. In such a case, by choosing r large enough (as a function of s) we have that $|\mathbf{Pr}[h(w) = a] - |\mathbb{F}|^{-1}| \leq |\mathbb{F}|^{-4s}$ for all $a \in \mathbb{F}$; and hence, if we set $\alpha_i = q_i(w'')$ for $1 \leq i \leq d'$ and $\alpha_0 = q_0(w'') - \Gamma''(w'')$, then

$$\mathbf{Pr}_{w' \in \mathbb{F}^{n'}}\left[\sum_{i=0}^{d'} \alpha_i h(w')^i = 0\right] = \mathbf{Pr}_{\beta \in \mathbb{F}}\left[\sum_{i=0}^{d'} \alpha_i \beta^i = 0\right] \pm |\mathbb{F}|^{-4s} \leq \frac{d'}{|\mathbb{F}|} + |\mathbb{F}|^{-4s},$$

where $\beta \in \mathbb{F}$ is a uniform field element. Combining these bounds, we have that

$$\mathbf{Pr}[\widetilde{f}(w'', h(w')) = \widetilde{g}(w'')] \leq \frac{d - d'd_0}{|\mathbb{F}|} + \left(1 - \frac{d - d'd_0}{|\mathbb{F}|}\right) \frac{d'}{|\mathbb{F}|} + |\mathbb{F}|^{-4s}$$

Recalling that $\mathbf{Pr}[\widetilde{f}(w'', h(w')) = \widetilde{g}(w'')] \geq \frac{e}{|\mathbb{F}|} + |\mathbb{F}|^{-2s}$, we obtain that

$$\frac{e}{|\mathbb{F}|} < \frac{d - d'd_0}{|\mathbb{F}|} + \frac{d'}{|\mathbb{F}|} = \frac{d - d'(d_0 - 1)}{|\mathbb{F}|}.$$

Thus, $d_0 - 1 < \frac{d-e}{d'} \leq d - e$ and hence $d_0 \leq d - e$ as claimed.

□

Chapter 7

Application of List Decoding in Randomness Extraction

7.1 Introduction

High-quality randomness is needed for a variety of applications. However, most physical sources are only weakly random. Moreover, such weak sources arise in cryptography when an adversary learns information about a uniformly random string. It is therefore natural and important to try to extract the usable randomness from a weak source. It is impossible to extract even one bit of randomness from a natural yet large enough class of sources using a single function [143]. There are two ways to counter this. One is to extract with the help of a small amount of randomness; this is called a seeded extractor [131]. Our focus is on the second way: to extract only from more structured sources (and not allow any auxiliary randomness). Such a function is called a *deterministic* (or seedless) extractor.

We now give a formal definition of extractors. In the following definition the term *source* simply refers to a random variable.

Definition 7.1.1. *A function* $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^m$ *is an* ϵ-extractor *for a family*

of sources $\mathcal{X}$ if for every $X \in \mathcal{X}$, the distribution $\mathrm{Ext}(X)$ *is ε-close in statistical (variation) distance to U_m. Here U_m denotes the uniform distribution on m bits.*

We measure the randomness in a source X using min-entropy.

Definition 7.1.2. *The* min-entropy *of a random variable X is*

$$H_\infty(X) = \min_{x \in \mathrm{supp}(X)} \log_2(1/\mathbf{Pr}[X = x]).$$

If $X \subseteq \{0,1\}^n$, we say that X has entropy rate $H_\infty(X)/n$.

The probabilistic method shows that if $|\mathcal{X}| \leq 2^{2 \cdot 9k}$, where k is the min-entropy of each source, then there exists a deterministic extractor for $\mathcal{X}$. Constructing such an extractor explicitly is a much harder challenge. One type of source for which deterministic extractors have been constructed is an *affine source* - a uniform distribution over an affine subspace of a vector space [63, 42, 52, 177, 117, 44]. In this paper, we explore generalizations of affine sources with more minimal structure. We show that an explicit deterministic extractor can be constructed for a broad generalization of affine sources that we call *additive sources*.

Remark 7.1.3. *Throughout the paper we often abuse notation and refer to a* set *X as a source. The source is actually the random variable uniformly distributed on the set X.*

Before presenting our general notion of an additive source, it will be instructive to look at two simpler natural generalizations of affine sources that are special cases of our notion. The first generalizes an affine source when viewed as the *image*

of a linear map. The second generalizes a linear source when viewed as the *kernel* of a linear map.

Generalized arithmetic progressions An affine subspace in $\mathbb{Z}_p^n$ may be viewed as the set of elements $V = \{a_1 \cdot t_1 + \ldots a_r \cdot t_r + b | t_1, \ldots, t_r \in \mathbb{Z}_p\}$ for some fixed $a_1, \ldots, a_r, b \in \mathbb{Z}_p^n$ such that $a_1, \ldots, a_r$ are linearly independent. One relaxation of this definition would be to not insist that $a_1, \ldots, a_r$ be linearly independent, and allow the t_i's to only range through a subset of $\mathbb{Z}_p$ of the form $\{0, \ldots, s-1\}$, rather than all of $\mathbb{Z}_p$. The result is exactly what is known as a *generalized arithmetic progression* (GAP). That is, a (r, s)-GAP is a set of the form

$$A = \{a_1 \cdot t_1 + \ldots a_r \cdot t_r + b | 0 \leq t_1, \ldots, t_r \leq s-1\}$$

for some fixed $a_1, \ldots, a_r, b \in \mathbb{Z}_p^n$ and $s \leq p$.

Bohr sets A linear subspace in $\mathbb{Z}_p^n$ may also be viewed as the set of elements $v \in \mathbb{Z}_p^n$ such that for all $i = 1, \ldots, d$, $L_i(v) = 0$, for some fixed linear functions $L_1, \ldots, L_d : \mathbb{Z}_p^n \to \mathbb{Z}_p$. A relaxation of this definition could be to look at the set of elements $v \in V$ such that $L_i(v)$ is 'close to zero' for every $i \in [d]$. We could define the *distance from zero* of an element $a \in \mathbb{Z}_p$ by looking at a as an integer in $\{0, \ldots, p-1\}$, and taking the minimum of the distances between $|p - a|$ and $|a - 0| = |a|$. Equivalently, we could define it as $\|a/p\|$ where $\|\cdot\|$ denotes the distance to the nearest integer. The resulting definition is what is known as a *Bohr set* in $\mathbb{Z}_p^n$. That is, a (d, ρ)-Bohr set is a set B of the form

$$B = \{v \in \mathbb{Z}_p^n : \|L_i(v)/p\| < \rho\}$$

for some fixed $0 < \rho < 1$ and linear functions $L_1, \ldots, L_d : \mathbb{Z}_p^n \to \mathbb{Z}_p$.

In fact, as opposed to subspaces GAPs and Bohr sets can be defined not just in $\mathbb{Z}_p^n$ but in any abelian group. See Definitions 7.2.3 and 7.2.6 for the definitions in a general abelian group.

We proceed to describe our general notion of an additive source.

7.1.1 Defining additive sources

Before defining an additive source, we give some intuition on the definition we chose and the pitfalls of other natural definitions.

We work in an abelian group G, which is usually $\mathbb{Z}_p$ or $\mathbb{Z}_p^n$ under addition. A first attempt at a minimal structure that generalizes subspaces is to require X to have small doubling: $|X + X| \leq C|X|$ for small $C > 1$. (Here $A + B$ denotes the set $\{a + b | a \in A, b \in B\}$.) The Cauchy-Davenport Theorem implies that for $A \subseteq \mathbb{Z}_p$, $|A + A| \geq \min\{2|A| - 1, p\}$. Kneser's theorem, which extends the Cauchy-Davenport theorem, implies the same is true for any $A \subseteq \mathbb{Z}_p^n$ that is not contained in a strict subgroup of $\mathbb{Z}_p^n$. So, for obtaining a large class of sources it makes sense to look at $C \geq 2$. However, even for $C = 2$ we get a class of sources for which deterministic extraction is impossible: For let $f : G \to \{0, 1\}$ be any such purported extractor. Then the uniform distribution on the larger of $f^{-1}(0)$ and $f^{-1}(1)$ gives a counterexample. If this seems artificial and one asks about smaller sets, we could start with any B such that $|B + B| \leq 2|B|$, such as an arithmetic progression, and then the larger of $f^{-1}(0) \cap B$ or $f^{-1}(1) \cap B$ gives a counterexample for $C = 4$. A similar attempt at a definition would be to lower bound the *additive energy* - a

quantity that measures how many sums in $X + X$ lead to the same value. However, this is also insufficient, as sets with small doubling have large energy.

In light of the above, we seek to impose an additional condition, besides small doubling. This extra condition involves the notion of symmetry sets from additive combinatorics. A symmetry set for a set $X \subseteq G$ with parameter $\gamma > 0$ is defined as

$$\mathrm{Sym}_\gamma(X) = \{g \in G : |X \cap (X + g)| \geq \gamma |X|\}.$$

In other words, an element is in $\mathrm{Sym}_\gamma(X)$ if it can be expressed as $x - x'$, for $x, x' \in X$, in at least $\gamma|X|$ ways. We shall be interested in the setting where γ is close to 1. The simplest examples of sets with large symmetry sets are subgroups and cosets of subgroups. Specifically, if X is a subgroup or a coset of a subgroup, then $\mathrm{Sym}_1(X) = X$.

We note that large symmetry sets don't imply small doubling. For example, if we start with a set Y with $\mathrm{Sym}_{1-\alpha}(Y)$ large, then we could choose a set T of size $2\alpha|Y|$ with large doubling, such as a Sidon set or random set, and set $X = Y \cup T$. Then $\mathrm{Sym}_{1-3\alpha}(X)$ is large but X has large doubling. Yet this counterexample isn't completely satisfactory, because for extraction it would suffice that whenever X has $\mathrm{Sym}_{1-\alpha}(X)$ large, there exists a large $X' \subseteq X$, $|X'| \geq (1-\varepsilon)|X|$, where X' has small doubling. We also give a counterexample to this weakened question. To give a counterexample with $p^{1/d}$ large symmetry sets, it's easiest to work in $\mathbb{Z}_p^d$. Pick a large-doubling set T in $\mathbb{Z}_p^{d-1}$, and let $X = T \times \mathbb{Z}_p$. Then $\mathrm{Sym}_1(X)$ contains $\mathbb{Z}_p$, but $X + X$ has size $\Theta(|T|^2 p) = \Theta(|X|^2/p)$. The same is true for large subsets of X. If we worked in $\mathbb{Z}_p$ instead, we could take a union of intervals, which would give

slightly weaker parameters.

Thus, we define an additive source to be (the uniform distribution on) a set that has small doubling and has a large symmetry set.

Definition 7.1.4 (Additive source)**.** *A set X in a finite abelian group $(G, +)$ is called an (α, β, τ)-additive source if $|X + X| \leq |X|^{1+\tau}$ and*

$$|\mathrm{Sym}_{1-\alpha}(X)| \geq |X|^{\beta}.$$

In Sections 7.3 and 7.4 we show that GAPs and Bohr sets in $\mathbb{Z}_p$ and $\mathbb{Z}_p^n$ are indeed captured by our definition of additive sources. As far as we know, these have not been studied in the extractor literature.

One can easily see that there are doubly exponentially many (α, β, τ)-additive sources for reasonably small α, τ and any constant $\beta < 1$. Due to this, there is no succinct representation of a general additive source, unlike affine sources. The only other natural family with doubly exponentially many sources is the family of independent sources.

7.1.2 Related work

We review relevant previous work. The first class of additive sources considered were bit fixing sources by Chor et al [47], and then by Kamp and Zuckerman [104] and Gabizon, Raz and Shaltiel in [65]. Next, in the more general case of affine sources, Bourgain obtained extractors for constant entropy rate [42] over $\mathbb{F}_2$, with improvements to slightly subconstant rate by Yehudayoff [177] and Li [117]. Recently Li [118] improved it to polylogarithmic min-entropy. In the case of large

fields, extractors for affine sources were given by Gabizon and Raz [63] and more recently by Bourgain, Dvir and Leeman [44]. DeVos and Gabizon [52] gave constructions interpolating between these extreme cases. Generalizations of affine sources have also been studied in the work of Dvir, Gabizon and Wigderson [55] and Ben-Sasson and Gabizon [18] where the authors look at polynomial sources and by Dvir [54] where varieties are considered. Special cases of affine sources have also been studied by Rao [137]. Gabizon and Shaltiel [64] constructed a weaker object called a disperser over large fields for affine sources. Ben-Sasson and Kopparty [19] constructed dispersers for affine sources with min-entropy $6n^{4/5}$ over $\mathbb{F}_2$. Shaltiel [148] improved on this and constructed a disperser for min-entropy $n^{o(1)}$ over $\mathbb{F}_2$.

7.1.3 Our results

In our main theorem for $\mathbb{Z}_p$ for p a large prime, we construct an extractor for additive sources for any constant entropy rate. More specifically, our construction works whenever p is large enough, and for (α, β, τ)-additive sources whenever α and τ are small enough. Specifically, for p an n-bit prime, we need $\alpha < 1/n$, and we extract about $\log(1/(\alpha n))$ bits. Thus, $\alpha = 1/\mathrm{poly}(n)$ leads to $\Omega(\log n)$ random bits whereas $\alpha = 1/p^{\gamma}$ leads to $\Omega(n)$ random bits. We now state our main theorem over $\mathbb{Z}_p$.

Theorem 7. *For every $\delta, \beta > 0$, there exists $\tau > 0$ and p_0 such that for all primes $p > p_0$ and $\alpha > 0$, the following holds. There is an explicit efficient ε-extractor* $\mathrm{Ext} : \mathbb{Z}_p \to \{0,1\}^m$*, for (α, β, τ)-additive sources of entropy rate δ in $\mathbb{Z}_p$ where* $\varepsilon = \left(3\alpha + p^{-\Omega_{\beta,\delta}(1)}\right) 2^{m/2} \log p$.

As a corollary, we obtain extractors for GAPs [1] for any constant entropy rate.

Corollary 8 (GAP Sources)**.** *For all $\delta > 0$, there exists c, p_0 such that for all primes $p \geq p_0$ the following holds. For all integers $r \geq c$, and all primes $p \geq c^{r/\delta}$ the following holds. There exists an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{Z}_p \to \{0,1\}^m$, for $(r, p^{\delta/r})$-GAP sources (of entropy rate δ) in $\mathbb{Z}_p$ where $\varepsilon = \left(\frac{3r}{p^{0.9\delta/r}} + p^{-1/2}\right) 2^{m/2} \log p$.*

Observe that the only restriction we put is that $r \geq c$ and $p \geq c^{r/\delta}$ which simply means that the side lengths of the GAP (that is, $p^{\delta/r}$) have to be larger than some fixed constant c and the dimension has to be larger than a fixed constant c. Thus, if we let r be a constant, then we can extract a constant fraction of the min entropy, that is $\Omega(\delta \log p)$ bits.

As another corollary, we obtain extractors for Bohr sources. We state it for constant ρ for simplicity. It can be easily generalized to any arbitrary ρ.

Corollary 9 (Bohr Sources)**.** *Let $\rho, \alpha > 0$ and $S \subseteq \mathbb{Z}_p$ with $|S| = d$ be arbitrary. Then for prime $p = \Omega\left(\left(\frac{d}{\alpha}\right)^d\right)$, there exists an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{Z}_p \to \{0,1\}^m$, for (d, ρ)-Bohr sources of entropy rate δ in $\mathbb{Z}_p$ where $\varepsilon = \left(3\alpha + p^{-\Omega(1)}\right) 2^{m/2} \log p$.*

Next, we construct an extractor for additive sources in $\mathbb{Z}_p^n$ for large enough p (polynomial in n) and any constant entropy rate, provided the source is sufficiently structured additively and satisfies a certain list decodability property. As a

[1] Technically, we prove it for proper GAPs (Definition 7.2.3). However, we prove a result for general GAPs by reducing to proper GAPs in Section 7.5.

corollary, we give an extractor for GAPs in $\mathbb{Z}_p^n$ and Bohr sets with constant entropy rate, provided they satisfy the list decodability property. We note that our extractor works for affine sources even though they may not satisfy the list decodability property. See Section 7.4.4.

Our extractors for GAPs and Bohr sets over $\mathbb{Z}_p$, and for Bohr sets over $\mathbb{Z}_p^n$, extract a linear number of bits with exponentially small error. We also show that all large sets (min-entropy rate close to 1), most sets (δ min-entropy rate for any $\delta > 0$) and most affine sources (δ min-entropy rate for any $\delta > 0$) satisfy the list decodability condition. See Remark 7.4.5, 7.4.6 and 7.4.7.

In the final two sections, we study special cases of additive sources. First, we give an extractor for one dimensional affine sources (lines) in $\mathbb{F}_q^n$ which requires only that $q > n$. This improves the results of Gabizon and Raz [63], which required $q = \Omega(n^2)$. Surprisingly, it even improves the non-explicit bound obtained via the probabilistic method of $q = \Omega(n \log n)$.

Theorem 10 (Extractors for lines)**.** *There is an explicit efficient ε-extractor* Ext : $\mathbb{F}_q^n \to \{0,1\}$ *for all line sources in $\mathbb{F}_q^n$ where $\varepsilon \leq 4(n/q)^{1/2}$.*

We then show the same extractor in fact works for 'partial lines' - i.e., arithmetic progressions in $\mathbb{Z}_p^n$.

Theorem 11 (Extractors for APs)**.** *There is an explicit efficient ε-extractor* Ext : $\mathbb{F}_p^n \to \{0,1\}^m$ *for all k-AP sources in $\mathbb{F}_p^n$ where $\varepsilon \leq 16 \log^2 p \sqrt{np} 2^{m/2}/k$.*

Therefore, if we have $k = p^{1/2+\delta}$ and $n < p^\delta$, then we can extract $\delta/2 \log p$ bits. Moreover, we show that the general framework of [63] for constructing extractors

for affine sources can be generalized to work for GAPs. See Theorem 7.5.10. As a corollary, we extend a result of DeVos and Gabizon [52] to obtain extractors for GAPs in $\mathbb{Z}_p^n$.

7.1.4 Techniques and Proof Overview

Extractors for additive sources in $\mathbb{Z}_p$.

For our proofs it will be convenient to define the notion of a *multiplicative source*. The definition simply corresponds to that of an additive source with multiplicative notation. Formally,

Definition 7.1.5 (Multiplicative source)**.** *Fix positive constants $0 < \alpha, \beta, \tau \leq 1$. Let Y be a subset of a finite abelian group $(G, \cdot)$. We define the set $\mathrm{Sym}_{1-\alpha}(Y) \subseteq G$ by*

$$\mathrm{Sym}_{1-\alpha}(Y) \triangleq \{g \in G : |Y \cap (Y \cdot g)| \geq (1-\alpha) \cdot |Y|\}.$$

We say that Y is an (α, β, τ)-multiplicative source if $|Y \cdot Y| \leq |Y|^{1+\tau}$ and $|\mathrm{Sym}_{1-\alpha}(Y)| \geq |Y|^{\beta}$.

Suppose X is an (α, β, τ)-additive source in $\mathbb{Z}_p$. Our extractor construction is as follows. We describe the construction in detail only for this class of sources. Let q be a large prime, $q = 1 \pmod{p}$ and g be a generator of $\mathbb{Z}_q^*$ of order p. Define $\mathrm{Ext}(x) \triangleq \sigma(g^x)$, where $\sigma : \mathbb{Z}_q \to \mathbb{Z}_m$ is the function from Lemma 7.2.8. Then, it is enough by Lemma 7.2.8, to show that $\left|\mathbb{E}_X e_p(a.g^X)\right|$ for all $a \neq 0$ is small. The analysis break down into two main steps:

Step 1: 'Encoding' X into a multiplicative source. As noted before, we fix a prime $q > p$ such that $q = 1 \pmod{p}$. For such q there exists an element $g \in \mathbb{Z}_q^*$ of order p. Fix such an element g and look at the map from $\mathbb{Z}_p$ to $\mathbb{Z}_q^*$ taking x to g^x. Let $Y \subseteq \mathbb{Z}_q^*$ be the image of X under this map. That is, $Y \triangleq \{g^x | x \in X\}$. As the subgroup generated by g in $\mathbb{Z}_q^*$ is isomorphic to $\mathbb{Z}_p$ we can show that Y is an (α, β', τ)-multiplicative source in $\mathbb{Z}_q^*$, where $\beta' \sim \beta$.

Step 2: Applying a character sum bound of Bourgain together with an 'average to worst-case reduction'. The advantage of the transition to a multiplicative source comes from a theorem of Bourgain that roughly says the following. Suppose Y is a subset of $\mathbb{Z}_q^*$ such that $|Y \cdot Y| \leq |Y|^{1+\tau}$ for appropriate $0 < \tau < 1$. Then, for most $a \in \mathbb{Z}_q$ the sum

$$\widehat{Y}(a) \triangleq \sum_{y \in Y} e_q(a \cdot y)$$

is small in absolute value. See Theorem 7.3.2 for a precise statement (The theorem does not directly correspond to the description here, and actually talks about the 't'th moment of additive characters over Y'.) If we knew that $|\widehat{Y}(a)|$ is small for *all* $a \in \mathbb{Z}_q^*$ rather than most $a \in \mathbb{Z}_q^*$, we could extract randomness from Y using the XOR lemma (Lemma 7.2.8). Our main insight is that when $\mathrm{Sym}_{1-\alpha}(Y)$ is large, we can indeed deduce that $|\widehat{Y}(a)|$ is small for all $a \in \mathbb{Z}_q^*$. We sketch why this is the case. Assume for contradiction that there is some $a \in \mathbb{Z}_q^*$ such that

$$|\widehat{Y}(a)| = \left| \sum_{y \in Y} e_q(a \cdot y) \right|$$

is large. Fix any $a' \in \mathrm{Sym}_{1-\alpha}(Y)$. As $|Y \cap a' \cdot Y| \geq (1-\alpha) \cdot |Y|$ and each summand is one in absolute value, the above sum will not change much if we sum over $a' \cdot Y$ rather than Y. That is, the sum

$$\sum_{y \in a' \cdot Y} e_q(a \cdot y)$$

must also be large in absolute value. But this sum is equal to

$$\sum_{y \in Y} e_q(a' \cdot a \cdot y) = \widehat{Y}(a' \cdot a).$$

Thus, $|\widehat{Y}(a' \cdot a)|$ is large for all $a' \in \mathrm{Sym}_{1-\alpha}(Y)$ - a contradiction as we know that $|\widehat{Y}(b)|$ is small for most $b \in \mathbb{Z}_q^*$.

Thus, for all $a \neq 0$, $\left|\mathbb{E}_X e_p(a.g^X)\right|$ is small. In summary, the extractor construction is $\mathrm{Ext}(x) \triangleq \sigma(g^x)$, where $\sigma : \mathbb{Z}_q \to \mathbb{Z}_m$ is the function from Lemma 7.2.8. See Section 7.3 for full details.

Extractors for additive sources in $\mathbb{Z}_p^n$.

Our construction over $\mathbb{Z}_p^n$ follows similar lines but is more involved. We give a sketch describing the same basic two steps. Let X be an (α, β, τ)-additive source in $\mathbb{Z}_p^n$.

Step 1: 'Encoding' X into a multiplicative source. We choose n distinct primes $q_1, \ldots, q_n$ such that for all $i \in [n]$, $q_i = 1 \pmod{p}$. Let g_i be an element of order p in $\mathbb{Z}_{q_i}^*$. Let $q = q_1 \cdots q_n$, and let $CRT : \prod_i \mathbb{Z}_{q_i} \to \mathbb{Z}_q$ be the 'Chinese remaindering map'.

We look at the map from $\mathbb{Z}_p^n$ to $\mathbb{Z}_q$ taking $(x_1, \ldots, x_n)$ to $CRT(g_1^{x_1}, \ldots, g_n^{x_n})$. Let Y be the image of X under this map. That is, $Y \triangleq \{CRT(g_1^{x_1}, \ldots, g_n^{x_n}) | (x_1, \ldots, x_n) \in X\}$.

We can show that Y is an (α, β', τ)-multiplicative source in $\mathbb{Z}_q^*$,[2] where $\beta' \sim \beta$ assuming $q = p^{O(1)}$. We show that we can indeed get $q = p^{O(1)}$ by observing that the proof of Linnik's Theorem implies that for large enough p, we can always find appropriate $q_1, \ldots, q_n$ that are all at most $p^{O(1)}$.

Step 2: Applying a character sum bound of Bourgain together with an 'average to worst-case reduction'. As in the case of $\mathbb{Z}_p$, we would now like to apply a theorem saying that for $Y \subseteq \mathbb{Z}_q^*$ such that $|Y \cdot Y| \leq |Y|^{1+\tau}$, $|\widehat{Y}(a)|$ is small for most $0 \neq a \in \mathbb{Z}_q$. The difference from the case of $\mathbb{Z}_p$ is that now we are dealing with a *composite* q. Bourgain indeed has such a theorem for the case of composite q. However, it requires an additional condition on Y apart from $|Y \cdot Y| \leq |Y|^{1+\tau}$. Roughly speaking, the condition is that if we look at elements of Y modulo a factor q_i of q, they are not too concentrated on any particular element of $\mathbb{Z}_{q_i}$. See Theorem 7.4.14 for a precise statement. We show that if X satisfies a certain 'list-decodability' condition, Y satisfies the condition required by Bourgain's theorem. For arbitrary $\gamma > 0$, we also show that a random source of min-entropy γn and a random affine source of min-entropy γn satisfy the list decodability condition with very high probability. Thus, our extractor does not work for all additive sources in $\mathbb{Z}_p^n$. The reduction from the statement about most $0 \neq a \in \mathbb{Z}_q$ to all is similar to

[2]Observe that since the vector $(g_1^{x_1}, \ldots, g_n^{x_n})$ is non-zero in all coordinates, the element $CRT(g_1^{x_1}, \ldots, g_n^{x_n})$ of $\mathbb{Z}_q$ is indeed in $\mathbb{Z}_q^*$.

the description in the case of $\mathbb{Z}_p$.

We show that for the case of affine sources, we do not need a list decodability condition on X. A potentially useful tool we develop for this is an XOR lemma that guarantees closeness to uniform under a weaker condition than usual. The usual setting, described for example by Rao [136], is when $N > M$ and for all nontrivial characters ψ on $\mathbb{Z}_N$, we have $\mathbb{E}_X[\psi(X)] \leq \varepsilon$. Then there's a simple map $\sigma : \mathbb{Z}_N \to \mathbb{Z}_M$ such that $\sigma(X)$ is close to uniform. We show that a similar result holds under the weaker assumption that $\mathbb{E}_X[\psi(X)] \leq \varepsilon$ only for characters ψ of the form $\psi(x) = e_n(a \cdot x)$ for $a \in \mathbb{Z}_N^*$, i.e., $(a, N) = 1$.

See Section 7.4 for full details.

Extractors for APs and GAPs.

For the case when the additive source is an AP or GAP in $\mathbb{Z}_p^n$, we give alternate constructions for a wider range of parameters.

For this, we generalize an approach introduced by Gabizon and Raz [63] and used by DeVos and Gabizon [52] for constructing extractors for affine sources. Their approach was to construct a polynomial $f : \mathbb{Z}_p^n \to \mathbb{Z}_p$ guaranteed to be non-constant on any k-dimensional affine subspace. Given such an f of degree d, the Weil bound (Theorem 7.2.12) can be used to construct an extractor for affine sources of dimension k when $p = \Omega(d^2)$. We show that the same approach works for GAPs: Suppose we can construct an explicit polynomial $f : \mathbb{Z}_p^n \to \mathbb{Z}_p$ of degree d that is non-constant *and of degree larger than one* when restricted to any affine subspace of dimension k. Then we can construct an extractor for GAPs in $\mathbb{Z}_p^n$ of dimension

$r \sim k$, assuming p is roughly $\Omega(d^2 \cdot \log^4 d)$. See Theorem 7.5.8 for a precise statement. Theorem 7.5.8 follows from a generalization of the Weil bound. Let us first recall the Weil bound of Theorem 7.2.12 says. Suppose we have a univariate polynomial f over $\mathbb{Z}_p$ of degree $d < \sqrt{p}$. Suppose ψ is a non trivial additive character of $\mathbb{Z}_p$. Then the character sum,

$$\left| \sum_{t \in \mathbb{Z}_p} \psi(f(t)) \right|$$

is small; more specifically, it is at most $d \cdot \sqrt{q}$. One may ask what happens when the same sum is taken only on the first s elements of $\mathbb{Z}_p$. Perhaps it is significantly larger than $d \cdot \sqrt{q}$ and becomes smaller only when running over the whole field? We show this is not the case. More precisely, for any $0 \leq s < p$ we have

$$\left| \sum_{0 \leq t \leq s-1} \psi(f(t)) \right| \leq 16 \log^2 p \cdot \sqrt{p} \cdot d$$

(see Lemma 7.5.11). The proof uses a combination of Theorem 7.2.12 and Fourier analysis. For example, a central step is to bound the Fourier coefficients of the function $\psi \circ f$ using the Weil bound.

See Section 7.5.2 for full details.

Extractors for lines over smaller fields. [63] used the approach mentioned above to construct extractors for line sources in $\mathbb{Z}_p^n$ over fields $\mathbb{Z}_p$ of size $p = \Omega(n^2)$. The main component in their construction was an explicit polynomial $f : \mathbb{Z}_p^n \to \mathbb{Z}_p$ of degree n that is non-constant when restricted to any affine line. We improve on this and construct a polynomial f of degree $O(\sqrt{n})$ that is non-constant on any affine line. As a result we get extractors for line sources in $\mathbb{Z}_p^n$ when $p = \Omega(n)$. We sketch the construction of f.

- The first step is to construct a polynomial $g : \mathbb{Z}_p^n \to \mathbb{Z}_p$ of degree n that is non-constant on any line, and moreover, has the following stronger property: The restriction of g to any affine line will have degree *exactly* n (rather than just *at most* n). We show that taking g to be a 'norm polynomial' insures this property.

- The second step is to partition the n coordinates into blocks of ascending size $1, 2, \ldots, \ell$ where $\ell = O(\sqrt{n})$. Now, let $g_i : \mathbb{Z}_p^i \to \mathbb{Z}_p$ denote the 'version' of the polynomial g when applied to a domain of i coordinates. We apply g_i to the i'th block. Note that the degree of the g_i's is at most $\deg(g_\ell) = \ell = O(\sqrt{n})$.

- Now we define f to be the sum of the g_i's when applied to the corresponding blocks. Note that $\deg(f) = O(\sqrt{n})$. We claim that f is non-constant on any affine line: Fix any affine line L, and fix the maximal $i \in [\ell]$ such that L is non-constant when restricted to the coordinates of the i'th block. The above-mentioned property of g guarantees that the 'g_i-summand of f' restricted to L will have degree i. All other summands will either be constant or of lower degree. Thus, f restricted to L is non-constant.

See Section 7.5 for full details.

7.1.5 Organization

In Section 7.2, we present basic definitions. In Section 7.3, we present our deterministic extractor for additive sources in $\mathbb{Z}_p$, and instantiate it in the case of GAPs and Bohr sets. In Section 7.4, we give our deterministic extractor for sources

in $\mathbb{Z}_p^n$ and again instantiate it in the case of GAPs and Bohr sets. In Section 7.5, we construct deterministic extractors for lines (1 dimensional affine spaces), partial lines in $\mathbb{Z}_p^n$ (APs) and further generalize to GAPs.

7.2 Definitions

In the following, p will denote a prime number. For $x \in \mathbb{R}$, $\|x\|$ denote the distance to the nearest integer. $e(x)$ denotes the complex number $e^{2\pi ix}$ and $e_m(x)$ denotes $e^{2\pi ix/m}$ for any positive integer m. To avoid clutter, e^y is written is $\exp(y)$.

7.2.1 Probability Distributions and Extractors

As mentioned earlier, a set X and a source X shall be used interchangeably where a source X denotes the uniform distribution on the set X.

7.2.2 Additive Combinatorics

We now state some standard terminology from additive combinatorics. We refer the reader to [158] for more details. In this section, let us fix a finite abelian group $(G, +)$.

Definition 7.2.1 (Representation function)**.** *Let A be a subset of G. For $g \in G$, we define*

$$rep_{A-A}(g) = |A \cap (A + g)|$$

which is the number of ways to represent g as a difference of two elements in A.

Definition 7.2.2 (Affine source and line source)**.** *A δ-affine source in $\mathbb{Z}_p^n$ is an affine source of dimension δn. A dimension 1 affine source is called a line source.*

Definition 7.2.3 (Generalized arithmetic progression)**.** *An (r, s)-Generalized arithmetic progression (or GAP for short) in G defined is a set of the form*

$$\left\{ b_0 + \sum_{i=1}^{r} a_i b_i : 0 \leq a_i \leq s-1 \right\}$$

for fixed elements $b_0, b_1, \ldots, b_r \in G$ (note that the a_i's are integers *rather than elements of G). We say that the GAP is* proper *if all the s^r sums are distinct. The dimension of the GAP is r.*

All GAPs are assumed to be proper in this paper unless mentioned otherwise. In fact, we will see in Section 7.5.2 how to handle general GAPs in $\mathbb{F}_p^n$.

Definition 7.2.4 (k-AP and k-line)**.** *An arithmetic progression of length k (or k-AP for short) is a $(1, k)$-GAP. A k-AP in $\mathbb{F}_q^n$ is also called a k-line.*

Definition 7.2.5 (k-HAP)**.** *A homogenous arithmetic progression of length k (or k-HAP for short) is a k-AP with $b_0 = 0$.*

Definition 7.2.6 (Bohr set)**.** *Let S be a set of characters of G and let $\rho > 0$. Then we define the Bohr set*

$$\mathbf{Bohr}(S, \rho) = \{x \in G : \max_{\xi \in S} |\xi(x)| < \rho\}$$

We call the ρ the radius and $|S|$ the rank of the Bohr set. We refer to a Bohr set of rank d and radius ρ as a (d, ρ)-Bohr set.

Bohr sets and GAPs are closely related. In particular, any Bohr set contains a large GAP with small dimension [141, Theorem 7.1].

We say that a Bohr set is *regular* if additionally,

$$\mathbf{Bohr}(S, \rho(1+\kappa)) \leq (1+100\kappa|S|)\mathbf{Bohr}(S,\rho)$$

whenever $\kappa < 1/100|S|$. Regular Bohr sets have the property that increasing the radius of the Bohr set by a little does not make the Bohr set very large. In fact, regular Bohr sets are ubiquitous [158], that is every Bohr set is "close" to a regular Bohr set. More precisely, for every S and ε, there is $\rho \in [\varepsilon, 2\varepsilon]$ such that $\mathbf{Bohr}(S,\rho)$ is regular. In this work, all Bohr sets will be regular Bohr sets.

When $G = \mathbb{Z}_p^n$, we know that the dual of G is isomorphic to G. Thus, in this case, we can consider $S \subseteq \mathbb{Z}_p^n$ and the Bohr set

$$\mathbf{Bohr}(S,\rho) = \{x \in \mathbb{Z}_p^n : \max_{\xi \in S} \left\| \frac{\xi \cdot x}{p} \right\| < \rho\}.$$

Here $\|\cdot\|$ denotes the distance to the nearest integer.

Note that if G is a vector space over $\mathbb{F}_q$, then every subspace of G is a Bohr set with radius $1/q$ and rank equal to the codimension of the subspace. Thus, Bohr sets are generalizations of subspaces and are substitutes for the latter when G has no proper subgroups (e.g., when G= $\mathbb{Z}_p$). Bohr sets can also be thought of as the inverse image of a cube in $\mathbb{C}^S$ (where $\mathbb{C}$ is the unit circle in $\mathbb{C}$) if one considers the map $x \mapsto (e_p(\xi \cdot x))_{\xi \in S}$. This is justified by the inequality $4\|\theta\| \leq |e(\theta)-1| \leq 2\pi\|\theta\|$.

7.2.3 Characters

Let $f : \mathbb{Z}_m \to \mathbb{C}$ be any function. Recall that, for $0 \leq j \leq m-1$, the Fourier coefficients of f are given by

$$\widehat{f}(j) = \frac{1}{m} \sum_{x \in \mathbb{Z}_m} f(x) \exp(-2\pi i j x/m).$$

It is well known that the set of functions $\{\exp(2\pi ijx/m)\}_{0\leq j\leq m-1}$ is an orthonormal basis for all complex functions defined on $\mathbb{Z}_m$, and that f can be expressed as

$$f(x) = \sum_{j=0}^{m-1} \widehat{f}(j) \exp(2\pi ijx/m).$$

Let us consider $f : \mathbb{Z}_m \to [0,1]$. Thus, Parseval's identity states that

$$\sum_{j=0}^{m-1} \left|\widehat{f}(j)\right|^2 = \frac{1}{m} \sum_{x\in\mathbb{Z}_m} f(x)^2 \leq 1.$$

Exponential/Character sums to extractors. Throughout the paper, ψ and χ denote additive and multiplicative characters respectively and ψ_0 and χ_0 denote the trivial additive and multiplicative characters respectively. We let $e_n(x)$ denote $e^{2\pi ix/n}$. We now state two lemmas that gives a black box construction of deterministic extractors from exponential/character sums. Note that we use the term exponential sum for additive characters and character sums for multiplicative characters.

The following lemma is for exponential sums.

Lemma 7.2.7. *Let $X \subseteq \mathbb{F}_p^n$. If $\left|\frac{1}{|X|}\sum_{x\in X}\psi(x)\right| < \varepsilon \ \forall \psi \neq \psi_0$, then there exists an efficient $\sigma : \mathbb{F}_p^n \to \mathbb{F}_p^m$ such that*

$$|\sigma(X) - U| < \varepsilon\sqrt{p^m}$$

We state a similar lemma that works for cyclic groups. A proof of this can be found in [135].

Lemma 7.2.8. *Let $X \subseteq \mathbb{Z}_N$. If $\left|\frac{1}{|X|}\sum_{x\in X}\psi(x)\right| < \varepsilon\ \forall \psi \neq \psi_0$, then there exists an efficient $\sigma : \mathbb{Z}_N \to \mathbb{Z}_M$ such that*

$$|\sigma(X) - U| < \varepsilon\sqrt{M}\log N + O(M/N)$$

The next lemma is for character sums.

Lemma 7.2.9. *Let $X \subseteq \mathbb{F}_{p^n}^*$. If $\left|\frac{1}{|X|}\sum_{x\in X}\chi(x)\right| < \varepsilon\ \forall \chi \neq \chi_0$, then there exists an efficient $\sigma : \mathbb{F}_{p^n}^* \to \mathbb{F}_{p^m}^*$ such that*

$$|\sigma(X) - U| < \varepsilon\sqrt{p^m}$$

Proof. Without loss of generality let us assume that m divides n. If not, we can always append 0's to increase the dimension by a factor of at most 2. We have the following standard claim.

Claim 7.2.10. *Let X be a distribution on G such that $|\mathbb{E}[\chi(X)]| \leq \varepsilon\ \forall \chi \neq \chi_0$. Then, X is $\varepsilon\sqrt{|G|}$ close to U.*

With the above claim, we define $\sigma : \mathbb{F}_{p^n}^* \to \mathbb{F}_{p^m}^*$ as $\sigma(x) = x^{\frac{p^n-1}{p^m-1}}$. Now,

Claim 7.2.11. *Given a nontrivial multiplicative character Ψ of $\mathbb{F}_{p^m}^*$, $\Psi \circ \sigma$ is a nontrivial multiplicative character of $\mathbb{F}_{p^n}^*$.*

Thus, by hypothesis, $\left|\frac{1}{|X|}\sum_{x\in X}\Psi\circ\sigma(x)\right| < \varepsilon$. Therefore, $\left|\mathbb{E}_{\sigma(X)}\Psi(\sigma(X))\right| < \varepsilon$. Thus, $\sigma(X)$ is $\varepsilon p^{m/2}$ close to U. □

The Riemann Hypothesis for curves over finite fields. In 1948 Weil [171] proved the celebrated *Riemann Hypothesis for curves over finite fields*. A consequence of Weil's result is a bound for exponential and character sums over low degree polynomials over a finite field. We state it below. The theorems can also be found in [144].

Theorem 7.2.12 (Weil's bound)**.** *Let ψ be a nontrivial additive character of $\mathbb{F}_q$. Let $f(t) \in \mathbb{F}_q[t]$ be a polynomial of degree m. Let $gcd(m, q) = 1$. Then*

$$\left| \sum_{t \in \mathbb{F}_q} \psi(f(t)) \right| \leq mq^{1/2}.$$

Theorem 7.2.13 (Weil's bound)**.** *Let χ be a nontrivial multiplicative character of $\mathbb{F}_q$ of order d. Let $f(t) \in \mathbb{F}_q[t]$ be a polynomial of degree m. Suppose that $f(t)$ is not of the form $cg(t)^d$ for any $c \in \mathbb{F}_q$ and $g(t) \in \mathbb{F}_q[t]$. Then*

$$\left| \sum_{t \in \mathbb{F}_q} \chi(f(t)) \right| \leq mq^{1/2}.$$

7.3 Extractors for additive sources in $\mathbb{Z}_p$

We now state our extractors for additive sources in $\mathbb{Z}_p$.

7.3.1 An extractor for additive sources

Our main theorem for $\mathbb{Z}_p$ (Theorem 7) follows from the following theorem.

Theorem 7.3.1. *Fix any $\delta > 0$ and positive constant C. There exists $p_0 \in \mathbb{N}$ such that for all primes $p \geq p_0$ the following holds. There is an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{Z}_p \to \{0,1\}^m$, for (α, β, τ)-additive sources of entropy rate δ in $\mathbb{Z}_p$*

where $\delta\beta \geq 2t\log_p(1/\alpha) + \delta/C$, $\varepsilon = 3\alpha 2^{m/2}\log p + O(2^m/p)$ *and* τ, t *are constants depending only on* δ *and* C.

Proof of Theorem 7. Let $C = 2/\beta$ and $\gamma = \frac{\beta\delta}{4t}$. Then the hypothesis of the above theorem is satisfied if $\alpha > p^{-\gamma}$. The $2^m/p$ term can now be dropped by assuming without loss of generality $\gamma < 1/2$. Now, since, any (α, β, τ)-additive source is an (α', β, τ)-additive source for $\alpha < \alpha'$, this finishes the proof. □

The above theorem follows from Lemma 7.3.3 and Lemma 7.2.8.

Before we state Lemma 7.3.3 and prove it, we state the following theorem.

Theorem 7.3.2 ([41, Theorem 1']). *For all* $Q \in \mathbb{Z}_+$, *there is* $\tau > 0$ *and* $t \in \mathbb{Z}_+$ *such that if* $H \subseteq \mathbb{F}_p^*$ *satisfies* $|H \cdot H| < |H|^{1+\tau}$, *then*

$$\frac{1}{p}\sum_{a=0}^{p-1}\left|\sum_{x\in H} e_p(ax)\right|^{2t} < |H|^{2t}\left(C_Q|H|^{-Q} + p^{-1+1/Q}\right)$$

Lemma 7.3.3. *There exists* $p_0 \in \mathbb{N}$ *such that for all primes* $p \geq p_0$ *the following holds. There exists an efficient* $f : \mathbb{Z}_p \to \mathbb{Z}_q$ *(for* $q = o(p^6)$*) such that if* $\delta > 0$ *is arbitrary and* C *is an arbitrary large constant, then there exist* $\tau(\delta, C) > 0$ *and* $t(\delta, C)$ *such that if*

- X *is an* (α, β, τ)*-additive set of entropy rate* δ *in* $(\mathbb{Z}_p, +)$,
- $\beta\delta \geq 2t\log_p(1/\alpha) + \delta/C$,

then, for all $\xi \in \mathbb{Z}_q \setminus \{0\}$,

$$\left|\sum_{x\in X} e_q(\xi f(x))\right| < 3\alpha|X|$$

.

Proof. Let q be the smallest prime such that $q = 1 \pmod{p}$. By Linnik's theorem, such a q exists and $q = O(p^{5.2})$. Let g be an element of $\mathbb{Z}_q^* \subset \mathbb{Z}_q$ such that $ord(g) = p$. Now define $f : \mathbb{Z}_p \to \mathbb{Z}_q$ as follows. Let $f(x) = g^x$. Let $Y = f(X)$. Let $S = Sym_{1-\alpha}(X)$.

Claim 7.3.4. *Y is a (α, β, τ)-multiplicative set of entropy rate $\delta/6$ in $(\mathbb{Z}_q^*, \times)$.*

Proof. We first note that $f : X \to Y$ is injective. To see that, suppose for $x, y \in \mathbb{Z}_p$ we have $g^x = g^y$. This implies $g^{x-y} = 1$. Since $\mathrm{ord}(g) = p$, we have $x \equiv y \pmod{p}$. Now, since f is injective, $|Y| \geq p^\delta \geq q^{\delta/6}$. Also, we claim that for each $a \in S$, $rep_{Y \cdot Y^{-1}}(f(a)) \geq (1-\alpha)|Y|$: This is because if $a \equiv x - x' \pmod{p}$ for $x, x' \in X$, then $g^a = g^{x-x'} = f(x)/f(x') \in Y \cdot Y^{-1}$, where the first equality again uses the fact that $\mathrm{ord}(g) = p$. So $|Y \cap (Y \cdot f(a))| = |X \cap (X + a)| \geq (1-\alpha)|Y|$. Now, observe that $|f(S)| \geq |X|^\beta = |Y|^\beta$. Finally, we show that $|Y \cdot Y| \leq |Y|^{1+\tau}$. This follows from the fact that for $a, b \in X$, $f(a) \cdot f(b) = f(a+b)$ and therefore $|Y \cdot Y| = |X + X|$. $\square$

We now continue with the proof. Let $Q = 6C/\delta$. Let $M = \max_{\xi \neq 0} \left| \sum_{y \in Y} e_q(\xi y) \right|$ and let ξ attain M.

Claim 7.3.5. $M < 3|Y|\alpha$.

Proof. Suppose $M \geq 3|Y|\alpha$. Consider any $\xi' \in f(S)$. Then,

$$\begin{aligned} \left| \sum_{y \in Y} e_q(\xi' \xi y) \right| &= \left| \sum_{y \in \xi' Y} e_q(\xi y) \right| \geq \left| \sum_{y \in Y} e_q(\xi y) \right| - 2(|Y| - |Y \cap \xi' Y|) \\ &\geq M - 2|Y|\alpha \geq |Y|\alpha. \end{aligned}$$

Since this lower bound holds for any $\xi' \in f(S)$, we have

$$|f(S)| \, |Y|^{2t} \, \alpha^{2t} \leq \sum_{\xi} \left| \sum_{y \in Y} e_q(\xi y) \right|^{2t}.$$

Therefore, since $|f(S)| \geq p^{\delta\beta}$, and Y satisfies the hypothesis of Theorem 7.3.2, we have

$$p^{\delta\beta} \alpha^{2t} \leq q \left(C_Q \, |Y|^{-Q} + q^{-1+1/Q} \right) < p^{\delta/C}$$

for large enough p. But this implies $\delta\beta - 2t \log_p(1/\alpha) < \delta/C$ which is a contradiction. Thus, we have $\max_{\xi \neq 0} \left| \sum_{y \in Y} e_q(\xi y) \right| < 3|Y|\alpha$. □

This implies

$$\left| \sum_{x \in X} e_q(\xi f(x)) \right| < 3|X|\alpha.$$

□

We proceed to formally show that GAPs and Bohr sets are indeed additive sources in $\mathbb{Z}_p$. We then use Theorem 7.3.1 to derive corollaries on these types of sources.

7.3.2 Application to GAPs and Bohr sets

We first show that a GAP source is an additive source with the appropriate parameters. We assume that the GAP is proper in this subsection.

Lemma 7.3.6. *For all $\varepsilon > 0$, there exists $c, n_0 \in \mathbb{N}$ such that for all prime $p \geq n_0$ the following holds. If $\delta \geq c/\log p$, then an $\left(r, p^{\delta}\right)$-GAP source is a $\left(r/p^{0.9\delta}, 0.1, \varepsilon\right)$-additive source of entropy rate δr in $(\mathbb{Z}_p, +)$.*

Proof. Let X be the $\left(p^\delta, r\right)$-GAP source defined by $X = \{b_0 + \sum_{i=1}^{r} a_i b_i : 0 \leq a_i \leq s-1\}$ where $s = p^\delta$ and let it be $(\alpha_0, \beta_0, \tau_0)$-additive. It is easy to see that the entropy rate is δr. The lemma now follows from a series of claims.

Claim 7.3.7. *$\tau_0 \leq \varepsilon$ for all $\varepsilon > 0$.*

Proof. Note that

$$X + X = \{2b_0 + \sum_{i=1}^{r} a_i b_i : 0 \leq a_i \leq 2s-2\}$$

Therefore, $|X + X| \leq 2^r s^r = 2^r |X|$ since X is a proper GAP. Now, $2^r < |X|^{\tau_0}$ iff $s^{\tau_0} > 2$ which is true for constant $\tau_0 = \varepsilon$ since $s = p^\delta \geq 2^c$. □

Claim 7.3.8. *$\alpha = r/p^{0.9\delta}$ and $\beta = 0.1$.*

Proof. Consider the set $S = \{b_0 + \sum_{i=1}^{r} a_i b_i : 0 \leq a_i < s^{0.1}\}$. Now fix an arbitrary $x \in S$. Then,

$$X \cap (X + x) \supseteq \{b_0 + \sum_{i=1}^{r} a_i b_i : s^{0.1} \leq a_i < s\}$$

Therefore, $|X \cap (X + x)| \geq (s - s^{0.1})^r = |X| \left(1 - 1/s^{0.9}\right)^r > |X| \left(1 - r/s^{0.9}\right)$. Also, we have $|S| \geq |X|^{0.1}$. This proves the claim. □

□

Note that the requirement of $\delta \geq c/\log p$ merely means that the sides of the GAP are $p^\delta = \Omega(1)$ in length.

Next, we show that a Bohr set is an additive source with the appropriate parameters. We will use the following lemma from [158] for the group G. As before let S be a set of frequencies of G.

Lemma 7.3.9 (Lemma 4.20 [158]). $|\mathbf{Bohr}(S,\rho)| \geq \rho^{|S|}|G|$ *and* $|\mathbf{Bohr}(S,2\rho)| \leq 4^{|S|}|\mathbf{Bohr}(S,\rho)|$.

We are now ready to prove our lemma about Bohr sets.

Lemma 7.3.10. *Let $\beta,\varepsilon,\delta,\rho > 0$ be arbitrary and $S \subseteq \widehat{G}$ be a set of frequencies. Let $B = \mathbf{Bohr}(S,\rho)$ in G where $d = |S|$. Let $0 \leq \kappa \leq \frac{1}{100d}$. A Bohr source is a $(100\kappa d,\beta,\varepsilon)$-additive source of entropy rate δ in G whenever $|G| \geq \max\left\{\left(\frac{4^{1/\varepsilon}}{\rho}\right)^d, \left(\frac{1}{\rho}\right)^{d/1-\delta}, \left(\frac{1}{\kappa\rho}\right)^{d/1-\beta}\right\}$.*

Proof. $|B| \geq \rho^d|G|$ by Lemma 7.3.9 and by the hypothesis, we have $|B| \geq |G|^\delta$.

To see that B has small doubling, observe that $B + B \subseteq \mathbf{Bohr}(S,2\rho)$ and therefore, using the fact $|\mathbf{Bohr}(S,2\rho)| \leq 4^d\mathbf{Bohr}(S,\rho)$ (Lemma 7.3.9) we have $|B+B| \leq 4^d|B| < |B|^{1+\varepsilon}$. The last inequality is true because $|B| \geq \rho^d|G|$ (Lemma 7.3.9) and $\rho^d|G| > 4^{d/\varepsilon}$ by hypothesis.

We now argue the presence of large symsets in B. Let $Y = \mathbf{Bohr}(S,\kappa\rho)$. Fix $y \in Y$. For any $x \in \mathbf{Bohr}(S,(1-\kappa)\rho)$, $x + y \in B$. Therefore,

$$|B \cap (y+B)| \geq |\mathbf{Bohr}(S,(1-\kappa)\rho)| \geq (1-100\kappa d)|B|$$

This is because we consider regular Bohr sets. Also, $|Y| \geq (\kappa\rho)^d|G| > |G|^\beta \geq |B|^\beta$ by the hypothesis and Lemma 7.3.9. This finishes the proof. □

GAP sources. We first restate our corollary for GAP sources.

Corollary 8. For all $\delta_0 > 0$, there exists $c, p_0 \in \mathbb{N}, \delta_0 > 0$ such that for all primes

$p \geq p_0$ the following holds. There exists an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{Z}_p \to \{0,1\}^m$, for (r, p^δ)-GAP sources (of entropy rate $\delta_0 = \delta r$) in $\mathbb{Z}_p$ where $p^\delta \geq c$, $r \geq C_{\delta_0}$ (where C_{δ_0} is a constant depending on δ_0 only) and $\varepsilon = 3\left(r/p^{0.9\delta} + p^{-1/2}\right) 2^{m/2} \log p$.

Proof. By Lemma 7.3.6, an (r, p^δ)-GAP source is a $\left(r/p^{0.9\delta}, 0.1, \tau\right)$-additive source of entropy rate δr in $\mathbb{Z}_p$ for all $\tau > 0$. We will apply Theorem 7.3.1 with $C = 20$ and $\delta' = \delta r$ which is a constant. Therefore, t and τ from the theorem conclusion are also constants. Note that we already have $s > 2^c$ by Lemma 7.3.6. Now, the second condition in Theorem 7.3.1 is equivalent to $0.1\delta r > 2t \log_p \left(p^{0.9\delta}/r\right) + \delta r/20$ is satisfied if $r \geq 36t$. We also drop the $2^m/p$ by putting a $p^{-1/2}$ term similar to Theorem 7. This finishes the proof. □

Note that the above extractor works for GAPs with sides as small as superconstant. It works as long as the total volume of the GAP exceeds $p^{\Omega(1)}$ which clearly improves upon a standard convex combination argument by extracting of each individual AP as that would need at least $p^{\Omega(1)}$ entropy along each side.

Bohr sources. We now restate our corollary for Bohr sources.

Corollary 9. Let $\rho, \alpha > 0$ and $S \subseteq \mathbb{Z}_p$ with $|S| = d$ be arbitrary. Then for prime $p = \Omega\left(\left(\frac{d}{\alpha}\right)^d\right)$, there exists an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{Z}_p \to \{0,1\}^m$, for (d, ρ)-Bohr sources of entropy rate δ in $\mathbb{Z}_p$ where $\varepsilon = \left(3\alpha + p^{-\Omega(1)}\right) 2^{m/2} \log p$.

Proof. By Lemma 7.3.10, any $\mathbf{Bohr}(S, \rho)$ is a $(100\kappa d, \beta, \varepsilon)$-additive source of entropy rate δ in $\mathbb{Z}_p$ whenever $p \geq \max\left\{\left(\frac{4^{1/\varepsilon}}{\rho}\right)^d, \left(\frac{1}{\rho}\right)^{d/1-\delta}, \left(\frac{1}{\kappa\rho}\right)^{d/1-\beta}\right\}$. The state-

ment follows from the lower bound on p and Theorem 7. $\square$

Note that the upper bound on δ is reasonable because as δ increases the Bohr structure keeps fading away. Hence we cannot extract from arbitrarily large Bohr sets.

7.4 Extractors for additive sources in $\mathbb{Z}_p^n$

We now state our extractors for additive sources in $\mathbb{Z}_p^n$.

7.4.1 GAPs and Bohr sets

We first show that a GAP source is an additive source with the appropriate parameters.

Lemma 7.4.1. *For all $\varepsilon > 0$, there exists $c, n_0 \in \mathbb{N}$ such that for all prime $p \geq n_0$ the following holds. If $\delta \geq (C/\log p)$, then an $\left(r = \mu n, p^\delta\right)$-GAP source is a $\left(\mu n/p^{0.9\delta}, 0.1, \varepsilon\right)$-additive source of entropy rate $\delta\mu$ in $(\mathbb{Z}_p, +)$.*

Proof. Let X be the $\left(r = \mu n, p^\delta\right)$-GAP source defined by $X = \{b_0 + \sum_{i=1}^{r} a_i b_i : 0 \leq a_i \leq s-1\}$ where $s = p^\delta$ and let it be $(\alpha_0, \beta_0, \tau_0)$-additive. It is easy to see that the entropy rate is $\mu\delta$. The lemma now follows from a series of claims.

Claim 7.4.2. *$\tau_0 \leq \varepsilon$ for all $\varepsilon > 0$.*

Proof. Note that

$$X + X = \{2b_0 + \sum_{i=1}^{r} a_i b_i : 0 \leq a_i \leq 2s-2\}$$

Therefore, $|X + X| \leq 2^r s^r = 2^r |X|$ since X is a proper GAP. Now, $2^r < |X|^{\tau_0}$ iff $s^{\tau_0} > 2$ which is true for constant $\tau_0 = \varepsilon$ since $s = p^\delta \geq 2^c$. $\square$

Claim 7.4.3. *$\alpha = r/p^{0.9\delta}$ and $\beta = 0.1$.*

Proof. Consider the set $S = \{b_0 + \sum_{i=1}^{r} a_i b_i : 0 \leq a_i < s^{0.1}\}$. Now fix an arbitrary $x \in S$. Then,

$$X \cap (X + x) \supseteq \{b_0 + \sum_{i=1}^{r} a_i b_i : s^{0.1} \leq a_i < s\}$$

Therefore, $|X \cap (X + x)| \geq (s - s^{0.1})^r = |X| (1 - 1/s^{0.9})^r > |X| (1 - r/s^{0.9})$. Also, we have $|S| \geq |X|^{0.1}$. This proves the claim. □

□

Note that the requirement of $\delta \geq (C/\log p)$ merely means that the sides of the GAP are $p^\delta = \Omega(1)$ in length.

We have already shown in Lemma 7.3.10 that Bohr sets are additive sources. We now proceed with the main theorem of this section.

7.4.2 Extractor for additive sources

We say that a set X is (r, B)-list decodable if for any arbitrary r indices $i_1, \cdots i_r$, $c_j \in \mathbb{Z}_p$ for $j \in [r]$, $\left|X_{x_{i_1}=c_1,\ldots x_{i_r}=c_r}\right| \leq B$. We now state the main theorem of this section.

Theorem 7.4.4. *There exists $p_0, L_0 \in \mathbb{N}$ such that for all $L \geq L_0$ and primes $p \geq p_0$ the following holds. Let $\gamma, \kappa > 0$ be arbitrary. There exists an efficient ε-extractor* $\mathrm{Ext} : \mathbb{Z}_p^n \to \{0,1\}^m$ *for $(\alpha, \kappa L/\delta, \tau)$-additive sources of entropy rate δ in $(\mathbb{Z}_p^n, +)$ where $n \leq p^{L-2}$, X is $(r, p^{-r\cdot\gamma\cdot L} \cdot |X|)$-list decodable for every $\tau n/L \leq r \leq n$ and $\varepsilon < (3\alpha + |X|^{-\tau})\, 2^{m/2} \log p^n + O(2^m/p^n)$ where τ is a constant depending on γ and κ.*

The following remarks show that the list decodability condition is not too restrictive.

Remark 7.4.5 (Min-entropy $(1-\varepsilon')n$ is list decodable)**.** *In fact, for high enough min-entropy, we can now eliminate the list decodability assumption altogether. Given L, choose $\gamma = 1/(4L)$, (and fix some $\kappa > 0$ as in Theorem 7.4.4). Now this fixing of γ and κ also fixes some $\tau > 0$. Denote $a = \tau n/L$. We claim the theorem can now be applied to* any *source of entropy $k = n - a/2$. Fix such a source X. For $r > a$, and any fixing of any r coordinates, the corresponding list will be of size at most p^{n-r}. We need to show that this is smaller than $p^{-L\gamma r}|X| = p^{-r/4+n-a/2}$ It can be checked that this is indeed the case*

$$n - r < -r/4 + n - a/2 \qquad iff \qquad (3/4)r > a/2$$

Remark 7.4.6 (Random set of min-entropy $\varepsilon' n$ is list decodable)**.** *A random set X of size $|X| > p^{2L\gamma n}$ satisfies the list decodability condition with high probability. To see this, fix $r > \tau n/L$, a set of indices S of size r, field values $c_1, \ldots, c_r$ for those indices and a subset W of the set of size $B = p^{-\gamma Lr}|X| + 1$. The probability that all of W has the property that the coordinates in S get values c_i's is $(1/p^r)^B$. A union bound over all W gives $\binom{|X|}{B}(1/p^r)^B < (|X|e/B)^B(1/p^r)^B << 1/p^{0.9rB}$ as γ is arbitrarily small. An outer round of union bound over each of the p^r settings of c_i's and S is too mild to boost up the error probability for large p.*

Remark 7.4.7 (Random affine source of min-entropy $\varepsilon' n$ is list decodable)**.** *Let $\gamma < 1/L$ be an arbitrary small constant. A subspace X of dimension $k > 2\gamma Ln$ defined by a random $k \times n$ matrix satisfies the list decodability condition. Indeed,*

let G be the random $k \times n$ matrix. We know that for any submatrix C of r columns in G, C has rank at least $\gamma r L$ with high probability. To see this, fix a subset of r column indices. Let $a = \gamma r L$. Note that $a < k/2$. Let C be the submatrix of G defined by the r columns. Then, $\mathbf{Pr}[\text{rank}(C) < a] < \binom{r}{a} p^{a-k}$. Here we are saying that some choice of the a columns of C will be linearly independent and then using the bound that a random $s \times t$ matrix has full rank with probability roughly at least $1 - p^{s-t}$ (for $s < t/2$). Continuing with the analysis, $\binom{r}{a} p^{a-k} < 2^r p^{-k/2} < 2^n p^{-k/2}$. Taking a union bound over the choice of r columns, we incur another factor of 2^n, and taking $p \geq 5^{2n/k}$ gives error at most $(4/5)^n$, finishing the proof. Let us continue with the proof. Fix r coordinates $i_1, \ldots, i_r$. Let $c_1, \ldots, c_r$ be r values in the field. Since the corresponding submatrix C has rank at least $\gamma r L$, the number of strings in X which are c_j in coordinate i_j, $j = 1, \ldots, r$, is at most $p^{-\gamma L r}|X|$. This satisfies the list decodability condition for all r.

The theorem follows from Lemma 7.4.18 and the following lemma.

Lemma 7.4.8. *There exists $p_0, L_0 \in \mathbb{N}$ such that for all $L \geq L_0$ and primes $p \geq p_0$ the following holds. There exists an efficient $f : \mathbb{Z}_p^n \to \mathbb{Z}_q$ (for $p^n < q < p^{Ln}$) such that if $\gamma, \kappa > 0$ are arbitrary, then there exists $\tau > 0$ such that if*

- *X is an $(\alpha, \frac{\kappa L}{\delta}, \tau)$-additive source of entropy rate δ in $(\mathbb{Z}_p^n, +)$*
- *$n \leq p^{L-2}$*
- *for every integer $\tau n/L \leq r \leq n$, X is $(r, p^{-r \cdot \gamma \cdot L} \cdot |X|)$-list decodable,*

Then, for all $\xi \in \mathbb{Z}_q \setminus \{0\}$,

$$\left|\sum_{x \in X} e_q(\xi f(x))\right| < 3 \max\{\alpha, 1/|X|^\tau\}|X|$$

Proof. Let $q_1, q_2, \cdots q_n$ be n distinct primes such that for all i, $q_i \equiv 1 \pmod{p}$. This is guaranteed by the following consequence of (the proof of) Linnik's theorem.

Claim 7.4.9. *There exist constants* $L_0 > 0$ *and* $n_0 \in \mathbb{N}$ *such that for all* $p \geq n_0, L \geq L_0$*, the size of the set*

$$\{q : q \equiv 1 \pmod{p}, q \leq p^L\}$$

is at least p^{L-2}*.*

Proof. Define

$$\theta(x, p) = \sum_{\substack{k \text{ prime}, k \leq x, \\ k \equiv 1 \pmod{p}}} \log k.$$

By [99, Corollary 18.8], there is a constant L_0 (known as *Linnik's constant*) such that for all p sufficiently large and $x \geq p^{L_0}$, we have

$$\theta(x, p) \geq \frac{Cx}{p^{1/2}\phi(p)} \geq \frac{Cx}{p^{3/2}}$$

for some constant C, where $\phi(n)$ is Euler's totient ϕ function. Let

$$\pi(x, p) = \sum_{\substack{k \text{ prime}, k \leq x, \\ k \equiv 1 \pmod{p}}} 1.$$

Then $\pi(x, p) \log x \geq \theta(x, p) \geq \frac{Cx}{p^{3/2}}$. Thus, $\pi(x, p) \geq \frac{Cx}{p^{3/2} \log x}$. If $x = p^L$ for $L \geq L_0$, then this is clearly $\geq p^{L-2}$. □

Thus, by the above, we have for all i, $q_i < p^L$. Also, let g_i generate the order p subgroup in $\mathbb{Z}_{q_i}^*$. Define two maps ϕ_1, ϕ_2 as follows. Let $\phi_1 : \mathbb{Z}_p^n \to \prod_{i\in[n]} \mathbb{Z}_{q_i}$ be defined by

$$\phi_1(x_1, x_2, \cdots x_n) = (g_1^{x_1}, \cdots g_n^{x_n})$$

and for $q = \prod_{i\in[n]} q_i$, let $\phi_2 : \prod_{i\in[n]} \mathbb{Z}_{q_i} \to \mathbb{Z}_q$ be defined by

$$\phi_2(y_1, \ldots y_n) = \sum_{i=1}^{n} y_i \frac{q}{q_i} \left[\left(\frac{q}{q_i} \right)^{-1} \right]_{q_i} \in \mathbb{Z}_q$$

where $[x^{-1}]_p$ is the inverse of x in $\mathbb{Z}_p^*$. Note that ϕ_2 is the Chinese remaindering map.

Define function f as follows.

$$f : \mathbb{Z}_p^n \to \mathbb{Z}_q$$

$$x \mapsto \phi_2 \circ \phi_1(x)$$

Let $Y = f(X)$. $|Y| \geq q^{\delta/L}$. In the following, define for q', q, $\pi_{q'} : \mathbb{Z}_q \to \mathbb{Z}_{q'}$ by $\pi_{q'}(x) = x \pmod{q'}$. Then, for $\xi \in \mathbb{Z}_{q'}$, we define $\pi_{q'}^{-1}(\xi) = \{x \in \mathbb{Z}_q : \pi_{q'}(x) = \xi\}$. We now have the following claim.

Claim 7.4.10. *If $q'|q$, and $q' > q^\tau, \xi \in \mathbb{Z}_{q'}$, then $|Y \cap \pi_{q'}^{-1}(\xi)| < (q')^{-\gamma}|Y|$.*

Proof. Without loss of generality, let $q' = \prod_{i=1}^{r} q_i$ for $r \leq n$. As $q' > q^\tau$, we have $p^{Lr} > q^\tau > p^{n\tau}$, since each $q_i \geq p$. Therefore, $r > \frac{n\tau}{L}$. Next, we need the following claim.

Claim 7.4.11. *Let q' be as above. Given any $\xi \in \mathbb{Z}_{q'}$, let $\xi_i = \xi \pmod{q_i}$ for $1 \leq i \leq r$. Then for $Y \subseteq \mathbb{Z}_q$, we have*

$$\left|Y \cap \pi_{q'}^{-1}(\xi)\right| = \left|X_{x_1=\log_{g_1}\xi_1,\ldots x_r=\log_{g_r}\xi_r}\right|$$

Proof. Note that $Y \cap \pi_{q'}^{-1}(\xi) = \{y \in Y : y \pmod{q'} = \xi\}$. The condition $y \pmod{q'} = \xi$ can be re-written as $y_i = \xi_i \pmod{q_i}$ as $q_i | q' | q$. Therefore, we have

$$\begin{aligned}
\{y \in Y : y \pmod{q'} = \xi\} &= \{y \in Y : y_i \pmod{q_i} = \xi_i,\ 1 \leq i \leq r\} \\
&= \{x \in X : g_i^{x_i} \pmod{q_i} = \xi_i,\ 1 \leq i \leq r\} \\
&= \{x \in X : x_i = \log_{g_i} \xi_i,\ 1 \leq i \leq r\}
\end{aligned}$$

□

Now, by the hypothesis, for $r \geq n\tau/L$, we have for any $c_1, \ldots, c_r$,

$$\begin{aligned}
|X_{x_1=c_1,\ldots x_r=c_r}| &\leq p^{-r\cdot\gamma\cdot L} \cdot |X| \\
&< (q')^{-\gamma} \cdot |Y|
\end{aligned}$$

as $p^{rL} > q'$. This finishes the proof. □

Next, we have the following.

Claim 7.4.12. $|Y.Y| < |Y|^{1+\tau}$

Proof. This follows because f is an one-one function from $\left(\mathbb{Z}_p^n, +\right)$ into $(\mathbb{Z}_q^*, *)$. □

Next, we have the following claim.

Claim 7.4.13. $\left|\mathrm{Sym}_{1-\alpha}(Y)\right| \geq q^{\kappa}$.

Proof. The proof follows because f is an one-one function from $\left(\mathbb{Z}_p^n, +\right)$ into $(\mathbb{Z}_q^*, *)$. It is similar to the proof of Claim 7.3.4. This would show that $\left|\mathrm{Sym}_{1-\alpha}(Y)\right| \geq |Y|^{\beta=\kappa L/\delta} = p^{n\kappa L} > q^{\kappa}$. □

We now need the following theorem due to Bourgain bounding the number of large Fourier coefficients.

Theorem 7.4.14 ([43, Corollary 3])**.** *Given $\gamma, \kappa > 0$, there is $\tau > 0$ such that the following holds. Let q be an arbitrary modulus and $H \subseteq \mathbb{Z}_q^*$ satisfy*

- *If $q'|q$, and $q' > q^{\tau}$, $\xi \in \mathbb{Z}_{q'}$, then $|H \cap \pi_{q'}^{-1}(\xi)| < (q')^{-\gamma}|H|$*
- *$|H.H| < |H|^{1+\tau}$.*

Then $\left|\{\xi \in \mathbb{Z}_q : \left|\sum_{x\in H} e_q(\xi x)\right| > |H|^{1-\tau}\}\right| < q^{\kappa}$.

Let $M = \max_{\xi\neq 0}\left|\sum_{y\in Y} e_q(\xi y)\right|$ and let ξ attain M. Let $S = \{x \in \mathbb{Z}_p^n : |rep_{X-X}(x)| > (1-\alpha)|X|\}$.

Claim 7.4.15. $M < 3|Y|\alpha$

Proof. Suppose $M \geq 3|Y|\alpha$. Consider any $\xi' \in f(S)$. Note that $|Y \cap \xi' Y| \geq$

$(1-\alpha)|Y|$. Then,

$$\begin{aligned}
\left|\sum_{y\in Y} e_q(\xi'\xi y)\right| &= \left|\sum_{y\in \xi' Y} e_q(\xi y)\right| \\
&\geq \left|\sum_{y\in Y} e_q(\xi y)\right| - 2(|Y| - |Y\cap \xi' Y|) \\
&\geq M - 2|Y|\alpha \\
&\geq |Y|\alpha \\
&\geq |Y|^{1-\tau}.
\end{aligned}$$

Since the above lower bound holds for any $\xi' \in f(S)$, we have a contradiction to Theorem 7.4.14 above as $|f(S)| = |X|^\beta = p^{\delta n(\kappa L/\delta)} > q^\kappa$. Thus, we have $\max_{\xi\neq 0}\left|\sum_{y\in Y} e_q(\xi y)\right| < 3|Y|\alpha$ □

This implies $\left|\sum_{x\in X} e_q(\xi f(x))\right| < 3\alpha|X|$, as desired. □

7.4.3 Application to GAPs and Bohr sets

We first state our corollary for GAP sources.

Corollary 7.4.16. *Let $C > 0$ be arbitrary. There exists $p_0, L_0 \in \mathbb{N}$ such that for all $L \geq L_0$ and primes $p \geq p_0$ the following holds. Let $\delta, \mu > 0$ be arbitrary. There exists an efficient ε-extractor $\mathrm{Ext} : \mathbb{Z}_p^n \to \{0,1\}^m$ for $(\mu n, p^\delta) - GAP$ sources (of entropy rate $\mu\delta$) in $\mathbb{Z}_p^n$ where $n \leq p^{L-2}$, X is $(\tau n/L, |X|^{1-1/C})$-list decodable and $\varepsilon < \left(3\frac{\mu n}{p^{0.9\delta}}\right) 2^{m/2}\log p^n + O(2^m/p^n)$ where $\tau < 1$ is a constant depending on $\delta \times \mu, L, C$.*

Proof. Choose $\kappa = \mu/10L$. By Lemma 7.4.1, a $(\mu n, p^{\delta}) - GAP$ is $(\mu n/p^{0.9\delta}, 0.1, \varepsilon)$-additive of entropy rate $\delta\mu$. To use Theorem 7.4.4, we need $\kappa L = 0.1\delta\mu$ which is true by the choice of κ. Now choose $\gamma = \delta\mu/CL$ for a large enough C. Then, for X that is $\tau(\delta\mu, C, L)n/L, |X|^{1-1/C})$-list decodable and $n \leq p^L - 2$, the hypothesis of Theorem 7.4.4 is satisfied and hence the statement follows. □

We now state our corollary for Bohr sets. As in the previous section, we state it for constant ρ and for $d = \mu n$ for simplicity.

Corollary 7.4.17. *Let $C, \rho, \alpha, \mu > 0$ be arbitrary. There exists $p_0, L_0 \in \mathbb{N}$ such that for all $L \geq L_0$ and primes $p \geq p_0$ the following holds. There exists an efficient ε-extractor $\mathrm{Ext} : \mathbb{Z}_p^n \to \{0,1\}^m$ for $(d = \mu n, \rho)$-Bohr sources in $\mathbb{Z}_p^n$ where $p \geq \max\{n^{1/(L-2)}, \Omega\left(\left(\frac{n}{\alpha}\right)^{\mu}\right)\}$, X is $(\tau n/L, |X|^{1-1/C})$-list decodable and $\varepsilon < (3\alpha + |X|^{-\tau})\, 2^{m/2} \log p^n + O(2^m/p^n)$ where $\tau < 1$ is an arbitrarily small constant depending on d, ρ and C.*

Proof. By Lemma 7.3.10, any $\mathbf{Bohr}(S, \rho)$ is a $(100\kappa d, \beta, \varepsilon)$-additive source of entropy rate δ in $\mathbb{Z}_p$ for $\kappa < 1/100d$ whenever $p^n \geq \max\left(\frac{4^{1/\varepsilon}}{\rho}\right)^d, \left(\frac{1}{\rho}\right)^{d/1-\delta}, \left(\frac{1}{\kappa\rho}\right)^{d/1-\beta}$. Now apply Theorem 7.4.4 and using the lower bound on p the conclusion follows. □

7.4.4 Application to affine sources and a new XOR lemma

We note that extractor for additive sources in $\mathbb{Z}_p^n$ presented above indeed works for arbitrary affine spaces of constant min-entropy without any condition on list decodability. Firstly we need a way of converting exponential sum bounds to extractors. This has been folklore and known as the Vazirani XOR lemma. However, the conditions required for that are too stringent for our character sum bounds and we need a different generalization of the XOR lemma which we state below.

Lemma 7.4.18. *Let $M < N$ be integers with M, N coprime and N be the product of n distinct primes all greater than p. Let $\sigma : \mathbb{Z}_N \to \mathbb{Z}_M$ be the function $\sigma(x) = x \mod M$. Let X be a distribution on $\mathbb{Z}_N$ with $|\mathbb{E}_X \psi(X)| \leq \varepsilon$ for every $\psi \in \mathbb{Z}_N^*$. Then,*

$$|\sigma(X) - U| = O\left((\varepsilon + n/p) \log N/M]\right)$$

The proof will perform a rather careful analysis of the traditional proof of the XOR lemma. We believe this might be of independent interest. Recently, Li obtained a better improvement to affine extractors over the binary field by getting the requirement down to polylogarithmic entropy [118].

7.5 Extractor for APs and GAPs in $\mathbb{F}_q^n$

We first focus our attention to the special case of line sources. We construct an extractor for line sources and later generalize to partial lines (or k-lines).

7.5.1 Extractor for lines in $\mathbb{F}_q^n$

As mentioned in the introduction, it becomes increasingly harder to construct an extractor for lines for small q (large n), since when n is large enough compared to q, we get a proof of non-existence by the density Hales-Jewett theorem. In this section, we shall focus on 1-bit extractors. Generalizations to more number of bits follows from the XOR lemma (Lemma 7.2.8). In the following, let q be power of p.

For the sake of completeness, we first show by a simple well known probabilistic argument, the existence of a 1-bit 0.1-extractor for lines sources in $\mathbb{F}_q^n$ as long as $q = \Omega(n \log n)$.

Lemma 7.5.1. *There exists a non-explicit* 0.1*-extractor* $f : \mathbb{F}_q^n \to \{0,1\}$ *for all line sources in* $\mathbb{F}_q^n$ *for* n *large enough as long as* $q > 200n \log n$.

Proof. Choose a random f such that for each x, $\mathbf{Pr}[f(x) = 0] = 1/2$. Fix an arbitrary source $X = \{a + tb : t \in \mathbb{F}_q\} \subseteq \mathbb{F}_q^n$. Recall that we view a set as a source which is uniform on the set. Let U denote the uniform distribution on $\{0,1\}$. For $i = 0, \ldots q-1$, let Y_i's be $0-1$ indicator random variables such that $Y_i = 1$ iff $f(a+ib) = 1$. We want to bound the event that $|f(X) - U| > 0.1$. This is equivalent to the event $\left|\frac{1}{q}\sum_i Y_i - 1/2\right| > 0.1$. Call the above event E_X. By a Chernoff bound, $\mathbf{Pr}[E_X] < 2\exp(-0.02q)$. By a union bound over all sources X, and noting that there are q^{2n} lines, $\mathbf{Pr}[f \text{ is not a 0.1 extractor}] < 2\exp(-0.02q)q^{2n} \leq 1$ by using the lower bound on q. □

Gabizon and Raz [63] achieved an extractor for $q = \Omega(n^2)$.

Theorem 7.5.2 ([63])**.** *There is an explicit efficient* ε*-extractor* $\text{Ext} : \mathbb{F}_q^n \to \{0,1\}$ *for all line sources in* $\mathbb{F}_q^n$ *where* $\varepsilon \leq n/\sqrt{q}$.

In this section, we construct our extractor which beats even the randomness argument and works for $q = \Omega(n)$.

The Main Theorem We state our main theorem of this subsection. As in the previous sections, the theorem follows from a lemma on exponential sums and the XOR lemma.

Theorem 10. There is an explicit efficient ε-extractor $\text{Ext} : \mathbb{F}_q^n \to \{0,1\}$ for all line sources in $\mathbb{F}_q^n$ where $\varepsilon \leq 4(n/q)^{1/2}$.

In order to construct our extractor, we shall be using *Norm Polynomials* p.272 of [119].

Definition 7.5.3 (Norm Polynomial)**.** *A norm polynomial $P \in F_q[r_1, \ldots r_k]$ is a homogeneous polynomial of degree k which satisfies for all $(c_1, c_2, \ldots, c_k) \in \mathbb{F}_q^k$, $P(c_1, \ldots c_k) = 0$ iff $c_1 = \ldots c_k = 0$.*

Construction of Norm Polynomials We follow the construction given in [119]. Let $\alpha_1, \ldots \alpha_k$ be a basis of $E = F_{q^k}$ over $\mathbb{F}_q$. Set

$$P(x_1, \ldots x_k) = \prod_{j=0}^{k-1} \left(\alpha_1^{q^j} x_1 + \ldots + \alpha_k^{q^j} x_k \right)$$

Since, the $\alpha_i^{q^j}$, $j = 0, 1, \ldots k-1$, are conjugates of α_i with respect to $\mathbb{F}_q$, the coefficients of N are in $\mathbb{F}_q$. Clearly, degree of N is d. Now let $(c_1, \ldots c_k) \in \mathbb{F}_q^n$. Then,

$$\begin{aligned} P(c_1, \ldots c_k) &= \prod_{j=0}^{k-1} \left(\alpha_1^{q^j} c_1 + \ldots + \alpha_k^{q^j} c_k \right) \\ &= \prod_{j=0}^{k-1} \left(\alpha_1 c_1 + \ldots + \alpha_k c_k \right)^{q^j} \\ &= \left(\alpha_1 c_1 + \ldots + \alpha_k c_k \right)^{\frac{q^k - 1}{q-1}} \end{aligned}$$

which is zero if and only if $\alpha_1 c_1 + \ldots + \alpha_k c_k = 0$, which is true iff $c_i = 0$ for all $1 \leq i \leq k$.

We now begin with the two main lemmas of this section. The first lemma is for additive characters and works for all q. The second lemma is for the quadratic multiplicative character for odd q.

Lemma 7.5.4. *There is an explicit efficient $f : \mathbb{F}_q^n \to \mathbb{F}_q$ such that the following holds. Let X be a line in $\mathbb{F}_q^n$. Then for any non trivial additive character ψ,*

$$\frac{1}{q}|\sum_{x \in X} \psi(f(x))| \leq 4(n/q)^{1/2}$$

Lemma 7.5.5. *Let q be odd. There is an explicit efficient $f : \mathbb{F}_q^n \to \mathbb{F}_q$ such that the following holds. Let X be a line in $\mathbb{F}_q^n$. Then for the multiplicative quadratic character χ_2 we have*

$$\frac{1}{q}|\sum_{x \in X} \chi_2(f(x))| \leq 4(n/q)^{1/2}$$

To extract more bits, we use the XOR lemma along with the theorem on additive characters. The general form is presented in the next subsection. Let us now focus on the problem of extracting 1 bit. When q is even, we use the trace function for the additive character which gives 1 bit. For odd q, we see that the quadratic character outputs 1 bit. Some care needs to be taken in this case. For a proof, see [63].

We start with the proof of Lemma 7.5.4.

Proof of Lemma 7.5.4. The construction is in two steps. First we define for an arbitrary subset of coordinates S, a polynomial Q_S and then partition the n coordinates carefully and apply a linear combination of the corresponding Q_S's.

Construction of Q_S Let $X = \{a + td : t \in \mathbb{F}_q\}$. For any subset of coordinates, say $S = \{x_1, \ldots, x_k\}$, we define $Q_S(t) = P(a_1 + td_1, \ldots a_k + td_k)$. Now, observe that

- Coefficient of t^k in $Q_S(t)$ is $P(d_1, \ldots d_k)$ which is zero iff $d_1 = \ldots d_k = 0$.

- On the other hand, we also have that if $d_1 = \ldots d_k = 0$, then $deg(Q_S) = 0$.

Combining the Q_S's Now, the construction is as follows. We partition the n coordinates into blocks of length $1, 2, 3, \ldots, d (d \leq 2\sqrt{n(1+1/p)})$ excluding multiples of p. Without loss of generality, we can assume that n is exactly partitioned in the increasing order as mentioned above. If not, we can always append all-zero coordinates and work in a dimension $< 2n$. (We adjust for this extra factor in the end. For now we assume n can be exactly partitioned) Let us call this family of subsets of coordinates $\mathbb{S}$. We let $f(t) = \sum_{S \in \mathbb{S}} Q_S(t)$. (We abuse notation and sometimes use $f(t)$ and $f(x)$ interchangeably with the obvious correspondence.) We now argue that this polynomial is nonzero whenever some d_i is nonzero. Now starting from the rightmost coordinate, we stop when we hit a nonzero d_i. All the blocks to its right will have degree 0 and all the ones to the left will have degree less than the degree of this block. So there is no cancellation. Thus, we always have a non zero polynomial of degree $d \leq 2\sqrt{2n(1+1/p)} < 4\sqrt{n}$ (taking the extra doubling of dimension into account) and we can apply Theorem 7.2.12 noting by the choice of the partition that $gcd(q, d) = 1$ (as d is never a multiple of p) to get

$$\frac{1}{q} |\sum_{t \in \mathbb{F}_q} \psi(f(t))| \leq 4(n/q)^{1/2}$$

□

Next, we prove Lemma 7.5.5.

Proof of Lemma 7.5.5. The construction is again in two steps. The first part is like in the previous proof but we state it for completeness. First we define for

an arbitrary subset of coordinates S, a polynomial Q_S and then partition the n coordinates carefully and apply a linear combination of the corresponding Q_S's.

Construction of Q_S Let $X = \{a + td : t \in \mathbb{F}_q\}$. For any subset of coordinates, say $S = \{x_1, \ldots, x_k\}$, we define $Q_S(t) = P(a_1 + td_1, \ldots a_k + td_k)$. Now, observe that

- Coefficient of t^k in $Q_S(t)$ is $P(d_1, \ldots d_k)$ which is zero iff $d_1 = \ldots d_k = 0$.
- On the other hand, we also have that if $d_1 = \ldots d_k = 0$, then $deg(Q_S) = 0$.

Combining the Q_S's Now, the construction is as follows. We partition the n coordinates into blocks of length $1, 3, \ldots, d (d \leq 2\sqrt{n})$, that is, excluding multiples of 2. Without loss of generality, we can assume that n is exactly partitioned in the increasing order as mentioned above. If not, we can always append all-zero coordinates and work in a dimension $< 2n$. (We adjust for this extra factor in the end. For now we assume n can be exactly partitioned) Let us call this family of subsets of coordinates $\mathbb{S}$. We let $f(t) = \sum_{S \in \mathbb{S}} Q_S(t)$. (We abuse notation and sometimes use $f(t)$ and $f(x)$ interchangeably with the obvious correspondence.) We now argue that this polynomial is nonzero whenever some d_i is nonzero. Now starting from the rightmost coordinate, we stop when we hit a nonzero d_i. All the blocks to its right will have degree 0 and all the ones to the left will have degree less than the degree of this block. So there is no cancellation. Thus, we always have a non zero polynomial of odd degree $d \leq 4\sqrt{n}$ (taking the extra doubling of dimension into account) and we can apply Theorem 7.2.13 noting by the choice of the partition

that the polynomial can never be a perfect square since it is of odd degree to get

$$\frac{1}{q}|\sum_{t\in\mathbb{F}_q}\chi_2(f(t))| \leq 4(n/q)^{1/2}$$

□

7.5.2 Extractors for APs and GAPs in $\mathbb{Z}_p^n$

We will build on the polynomial obtained in the previous subsection to get an extractor for APs and GAPs. As we will use field operations, it will be convenient to use the notation $\mathbb{F}_p^n$ rather than $\mathbb{Z}_p^n$. Fix integers r, s with $1 \leq s \leq p-1$. For $a_1, \ldots, a_r, b \in \mathbb{F}_p^n$ we denote by $G_{a_1,\ldots,a_r,b}$ the (r,s)-GAP $G_{a_1,\ldots,a_r,b} \triangleq \{\sum_{i=1}^r a_i \cdot t_i + b : 0 \leq t_i \leq s-1\}$.

It will be convenient to look at GAPs where the a_i's are linearly independent.

Definition 7.5.6. *For $a_1, \ldots, a_r, b \in Z_p^n$, we say the (r,s)-GAP $G_{a_1,\ldots,a_r,b}$ is* independent *if $a_1, \ldots, a_r$ are linearly independent in $\mathbb{Z}_p^n$.*

Claim 7.5.7. *An (r,s)-GAP in $\mathbb{F}_p^n$ can be written as a union of (k,s)-GAPs in $\mathbb{F}_p^n$ which are independent, for $k \geq r \cdot \log s / \log p$.*

Proof. Fix an (r,s)-GAP $G_{a_1,\ldots,a_r,b}$. Let k be the dimension of the $\mathbb{F}_p$-linear span of $\{a_1, \ldots, a_r\}$. We have

$$p^k \geq s^r \rightarrow k \geq r \cdot \log s / \log p.$$

Assume, w.l.o.g., that $a_1 \ldots, a_k$ are linearly independent. We can write $G_{a_1,\ldots,a_r,b}$ as a union of GAPs $G_{a_1,\ldots,a_k,b'}$ where b' will range over the values $\{a_{k+1}\cdot t_{k+1} + \ldots + a_r \cdot t_r + b : 0 \leq t_i \leq s-1\}$. □

[63] and [52] used polynomials that are non-constant over subspaces of a certain dimension together with Weil bounds to construct affine extractors. We show that such polynomials are sufficient for the more general goal of constructing extractors for GAPs. For a polynomial $f : \mathbb{F}_p^n \to \mathbb{F}_p$, and $a_1, \ldots, a_k, b \in \mathbb{F}_p^n$, we denote by $f|_{a_1,\ldots,a_k,b}$ the polynomial $f|_{a_1,\ldots,a_k,b}(t_1, \ldots, t_k) \triangleq f(a_1 \cdot t_1 + \ldots + a_k \cdot t_k + b)$. We first prove the following theorem.

Theorem 7.5.8. *Fix integers r, s, d with $d, s < p$. Fix integer $k \leq \min\{1, r \cdot \log s / \log p\}$. Suppose we are given an efficiently computable polynomial $f : \mathbb{F}_p^n \to \mathbb{F}_p$ such that for all $a_1, \ldots, a_k, b \in \mathbb{F}_p^n$, where $a_1, \ldots, a_k$ are linearly independent, $f|_{a_1,\ldots,a_k,b}$ is non-constant of degree $1 < d' \leq d$.*

Then we can construct an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{F}_p^n \to \mathcal{B}^m$ for (r, s)-GAP sources where

$$\varepsilon \leq (4 \log p \cdot \sqrt{p} + 1) \cdot d/s \cdot 2^{m/2} + 2^m/p.$$

Remark 7.5.9. *In the case $r = 1$ a d/s factor can be taken off from the ε. This will be evident in the proof.*

Once we have this, the two main theorems follow immediately.

Theorem 11. There is an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{F}_p^n \to \{0,1\}^m$ for all k-AP sources in $\mathbb{F}_p^n$ where $\varepsilon \leq 16 \log^2 p \sqrt{np} 2^{m/2}/k$.

Proof. Plugging the polynomial from Lemma 7.5.4 that has degree $4\sqrt{n}$ and is non-constant on affine subspaces of dimension 1 completes the proof. The $2^m/p$ factor can be dropped as it will be dominated by the first term. □

Theorem 7.5.10 (Extractors for GAPs)**.** *Fix integers r, s with $s < p$.*

Then we can construct an explicit efficient ε-extractor $\mathrm{Ext} : \mathbb{F}_p^n \to \mathcal{B}^m$ *for (r, s)-GAP sources where*

$$\varepsilon \leq (34 \log^3 p \cdot \sqrt{p}) \cdot n/(r \cdot \log s \cdot s) \cdot 2^{m/2} + 2^m/p.$$

In particular, when $p = \Omega((r \cdot s/n)^2)$ we can output one bit with constant error.

Proof. DeVos and Gabizon ([52], Theorem 7) construct an explicit function $f : \mathbb{F}_p^n \to \mathbb{F}_p$ that is non-constant of degree $1 < d < 2n/k$ when restricted to any affine subspace of dimension k. Plugging this function into Theorem 7.5.8 finishes the proof. □

Let us now prove Theorem 7.5.8. The main ingredient in the theorem's proof is the following lemma that generalizes the Weil bound for exponential sums (Theorem 7.2.12) to the case where the sum ranges only over an AP, rather than the whole field.

Lemma 7.5.11. *Let $f \in \mathbb{F}_p[t]$ be a polynomial of degree $1 < d < p$. Let X be an s-AP. Let ψ be a non trivial additive character of $\mathbb{F}_p$. Then, for any integer $0 < s \leq p$,*

$$\left| \sum_{t \in X} \psi(f(t)) \right| \leq 4 \log p \cdot \sqrt{p} \cdot d.$$

Using the XOR lemma, Lemma 7.5.11 implies the following.

Corollary 7.5.12. *Let $f \in \mathbb{F}_p[t]$ be a polynomial of degree $1 < d < p$. For any integer $0 < s \leq p$, let X be an s-AP source. Let $\sigma : \mathbb{F}_p^n \to \mathbb{F}_p^m$ be the function from Lemma 7.2.7. Then, $|\sigma(X) - U| < \varepsilon$ for $\varepsilon = 4 \log p \cdot \sqrt{p} \cdot d \cdot p^{m/2}$.*

We prove Theorem 7.5.8 given the corollary.

Proof. Let $\sigma : \mathbb{F}_p^n \to \mathbb{F}_p^m$ be the function from Lemma 7.2.7. Fix an (s,k)-independent GAP $X = G_{a_1,\ldots,a_k,b}$.

Claim 7.5.7 implies it is enough to construct an extractor for such GAPs.

We know that $g \triangleq f|_{a_1,\ldots,a_k,b}$ is non-constant of degree d' where $1 < d' \leq d < p$. Assume w.l.o.g. that t_1 appears in g with degree greater than 1. Denote by a the maximal degree that t_1 has in g. Write g as a polynomial in t_1 whose coefficients are polynomials in $t_2, \ldots, t_k$. Look at the coefficient $g_a(t_2, \ldots, t_k)$ of t_1^a in g. From the Schwartz-Zippel Lemma $g_a(t_2, \ldots, t_k) = 0$ with probability at most $d'/s \leq d/s$ when choosing $t_2, \ldots, t_k$ uniformly in $\{0, \ldots, s-1\}^{k-1}$. Note that X can be viewed as a convex combination of the s-APs $X_{t_2,\ldots,t_k} \triangleq \{a_1 t_1 + a_2 t_2 + \ldots + a_k t_k + b | 0 \leq t_1 \leq s-1\}$. For $t_2, \ldots, t_k$ such that $g_a(t_2, \ldots, t_k) \neq 0$, it follows that the polynomial $g_{t_2,\ldots,t_k}(t) \triangleq g(t, t_2, \ldots, t_k)$ is non-constant of degree larger than 1 and at most d. Therefore, from Corollary 7.5.12 we have $|\sigma(X_{t_2,\ldots,t_k}) - U| \leq 4 \log p \cdot \sqrt{p} \cdot d \cdot p^{m/2}$. $\square$

We proceed with the proof of Lemma 7.5.11.

Proof. (of Lemma 7.5.11)

The proof combines the Weil bound with Fourier analysis. It is based on two claims. The first uses the Weil bound to bound the Fourier coefficients of f *composed with an additive character.*

Claim 7.5.13. *Let ψ be the non trivial additive character from the lemma statement. Then for all $\xi \in \mathbb{F}_p$, we have $\left|\widehat{\psi \circ f}(\xi)\right| \leq \sqrt{d/p}$.*

Proof. Suppose $\psi(x) = e_p(a \cdot x)$ for some $a \in \mathbb{F}_p$. We have

$$\begin{aligned} & \left|\widehat{\psi \circ f}(\xi)\right| \\ = \; & 1/p \left| \sum_{t \in \mathbb{F}_p} e_p(a \cdot f(t)) \cdot e_p(-\xi t)) \right| \\ = \; & 1/p \left| \sum_{t \in \mathbb{F}_p} e_p(a^{-1}(f(t) - a \cdot \xi \cdot t)) \right| \\ \leq \; & (d/p)^{1/2} \end{aligned}$$

The last line follows from Weil bound and by observing two things: The first is that the sum in the line before is an exponential sum with the character $\phi'(x) = e_p(a^{-1} \cdot x)$ on the polynomial $f'(t) \triangleq f(t) - \xi\psi^{-1}t$. The second is that $f'(t)$ is also a non-constant polynomial of degree $d < p$ so the Weil bound can be used. □

Next, we need the following claim upper bounding the L_1 Fourier norm of a set related to s-APs. Let $A = \{0, 1, \ldots s-1\}$. Denote by $A(x)$ the indicator set of A.

Claim 7.5.14. $\sum_{0 \leq j \leq p-1} \left|\hat{A}(j)\right| \leq 4 \log p$

Proof. Note that

$$\begin{aligned} \hat{A}(j) &= 1/p \sum_{i \in A} e(ji) \\ &= 1/p \frac{e(jk) - 1}{e(j) - 1} \end{aligned}$$

Now noting that $|e(\theta) - 1| \geq 4\{\theta\}$ we have $\left|\hat{A}(j)\right| \leq \frac{1}{2p\{j/p\}}$

We now turn to computing the L_1 Fourier norm of A.

$$\begin{aligned}\sum_{j\in\mathbb{F}_p}\left|\hat{A}(j)\right| &= \sum_{j\leq p/2}\left|\hat{A}(j)\right| + \sum_{j>p/2}\left|\hat{A}(j)\right| \\ &\leq \sum_{j\leq p/2} 1/(2j) + \sum_{j>p/2} 1/2(p-j) \\ &\leq 4\log p\end{aligned}$$

□

With the above two claims in place, we now turn to proving the lemma.

$$\begin{aligned}\left|\sum_{0\leq t\leq s-1}\psi(f(t))\right| &= \left|\sum_{t\in\mathbb{F}_p} A(t)\psi(f(t))\right| \\ &= p\left|\sum_{\xi\in\mathbb{F}_p}\overline{\widehat{A}(\xi)}\widehat{\psi\circ f}(\xi)\right| \\ &\leq 4\log p\cdot\sqrt{p}\cdot d.\end{aligned}$$

This finishes the proof. □

Chapter 8

Conclusion

8.1 Error correcting codes

Error correcting codes are designed to tackle the problem of robust transmission of data through noisy channels. In this dissertation, we answer certain open questions about locally decodable and list decodable codes.

Locally decodable codes. A central research question, which is far from being solved, is understanding the best possible 'stretch' of an LDC with a constant number of queries. That is, how large N has to be as a function of K. The best family of LDCs in the constant query regime are Matching Vector (MV) codes. Our main contribution in this area is a lower bound on the $s$tretch of MV codes [32]. The first result proves a quadratic lower bound ($N = \Omega(K^2)$) which resolves a conjecture raised in [56]. The result holds even when the number of queries is not constant. The second result states that under a well known conjecture from additive combinatorics, one needs a super polynomial stretch ($N = K^{\log \log K}$), thus ruling out efficient MV codes for constant number of queries. An open problem is to achieve the superpolynomial lower bound without relying on any conjectures. In fact, any

unconditional super quadratic lower bound would be interesting.

The best constructions of MV codes come from constructions of a particular type of polynomial called an OR polynomial. Smaller degree OR polynomials lead to better MV codes. We show a barrier why existing techniques have failed to improve the degree bounds of OR polynomials for more than a decade. In fact, the barrier we discover, that is, nonclassical polynomials, applies to many other fundamental problems in complexity theory like correlation bounds and lower bounds. It would be interesting to find out other applications in computer science where nonclassical polynomials are a barrier to improving bounds in longstanding open problems.

List decodable codes. The concept of *list decoding* was introduced by Elias [60] and Wozencraft [176] to decode *error correcting codes* beyond half the minimum distance. The objective of list decoding is to output all the codewords within a specified radius around the received word. Despite so much progress, the largest radius up to which list decoding is tractable (the *list decoding radius*) is still a fundamental open problem even for well studied codes like Reed-Solomon (univariate polynomials) and Reed-Muller codes (multivariate polynomials). The goal of this work is to analyze Reed-Muller codes over small fields $\mathbb{F}$ and small degree d. The list decoding radius was conjectured to *approach* the minimum distance of the code. See Chapter 4 for a precise definition. This was proved for the $d = 1$ case [67, 68], the $\mathbb{F} = \mathbb{F}_2$ case [74] and the $d = 2$ case [73]. It was conjectured [74] to be true for all fixed fields and fixed degree. Our main contribution is a positive resolution of the above conjecture for fixed prime fields [35]. Moreover, we give a tight bound on the weight distribution of generalized Reed Muller codes over prime fields. This

is a fundamental problem in coding theory; see Research Problem (15.1) in [124]. In follow up work, we extend this to large fields and prove that the list decoding radius equals the minimum distance for fixed degree and all prime fields [36]. This is a consequence of a theorem about pseudorandomness of polynomials which is discussed in more detail later. An interesting open problem is to extend this to all degrees and to the setting of Reed Solomon codes.

Applications of List Decoding to Randomness Extraction. High-quality randomness is needed for a variety of applications. However, most physical sources are only weakly random. It is therefore natural and important to try to extract the usable randomness from a weak source. It is impossible to extract even one bit of randomness from a natural yet large enough class of sources using a single function [143]. One way to counter this is to extract only from more structured sources (and not allow any auxiliary randomness). Such a function is called a *deterministic* (or seedless) extractor. Our main contribution is to extract randomness from a very general class of *additive* sources which satisfy a certain list decodability property [33]. Our work generalizes many existing results. An open problem here is to improve the number of bits output in such cases.

8.2 Polynomials and computation.

Recall that f is *equidistributed* if f takes every value in $\mathbb{F}$ with the same frequency, else it is *biased*. We say that f is *low rank* if it can be expressed as a composition of bounded number of lower degree polynomials and high rank otherwise.

The dichotomy of bias versus low rank is universal and has seen applications in mathematics and computer science including graph theory, number theory, ergodic theory, discrete geometry, property testing, complexity and algorithms. The first result dates back to the breakthrough resolution of the Riemann hypothesis over function fields by Weil in 1948 (the univariate setting [172]) and by Deligne in 1975 (the multivariate setting [51]). As a consequence they show if a low degree polynomial $f : \mathbb{F}^n \to \mathbb{F}$ has bias at least $1/\sqrt{p}$, then $f(x) = \Gamma(h(x))$ for some lower degree polynomial $h(x)$. In fact, they even show what Γ looks like. A natural question to ask is what happens if the bias is not so high, but some arbitrary inverse polynomial in the field size, say $1/p^{10}$.

Green and Tao [80] and Kauffman and Lovett[108] in 2008 gave a partial answer to this in the following sense. They showed that if the bias of f is at least $1/p^k$, then $f(x) = \Gamma(h_1(x), \ldots, h_c(x))$ where $c = c(d, k, p)$, d is the degree of the polynomial and h_i's are lower degree polynomials. This is a partial answer because c is dependent on p and d. In fact the dependence is Ackermann in the above parameters. This makes the result trivial for any reasonably growing field.

We have shown in this dissertation that $c = c(d, k)$. That is, we remove any kind of dependence on p. Any type of bias versus low rank theorem has seen lot of applications in the areas highlighted above. For instance, we resolve the problem of list decoding radius of fixed degree Reed Muller codes and also give a very tight bound on their asymptotic weight distribution, an open problem since the 80's. Among other applications, we improve bounds in a suite of problems in effective algebraic geometry including Hilbert's Nullstellensatz, radical membership

and counting points in rational varieties. Currently all our improvements work as long as d is fixed. The positive thing however is that there is no longer any restriction on field size.

A very nice open problem is to make c independent of d or have better dependence on d. This would translate to the above improvements to hold even for growing degrees.

Bibliography

[1] Leonard M. Adleman and Ming-Deh A. Huang. Counting rational points on curves and abelian varieties over finite fields. In *Proceedings of the Second International Symposium on Algorithmic Number Theory*, ANTS-II, pages 1–16, London, UK, UK, 1996. Springer-Verlag.

[2] A. Akavia, S. Goldwasser, and S. Safra. Proving hard-core predicates using list decoding. In *Proc.* 44^{th} *IEEE Symposium on Foundations of Computer Science (FOCS'03)*, 2003.

[3] Noga Alon, Oded Goldreich, Johan Hastad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992.

[4] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Trans. Inform. Theory*, 51(11):4032–4039, 2005.

[5] Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23:365–426, 2003.

[6] James Ax. Zeros of polynomials over finite fields. *American Journal of Mathematics*, 86:255–261, 1964.

[7] Laszlo Babai, Lance Fortnow, Leonid Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *23rd ACM Symposium on Theory of Computing (STOC)*, pages 21–31, 1991.

[8] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.

[9] Laszlo Babai and Peter Frankl. *Linear algebra methods in combinatorics.* 1998.

[10] Antal Balog and Endre Szemerédi. A statistical theorem of set addition. *Combinatorica*, 14(3):263–268, 1994.

[11] D. Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *Journal of Computer and System Sciences*, 38, 1989.

[12] D.A. Barrington and G. Tardos. A lower bound on the mod 6 degree of the or function. *computational complexity*, 7(2):99–108, 1998.

[13] David A. Barrington, Richard Beigel, and Steven Rudich. Representing Boolean functions as polynomials modulo composite numbers. *Computational Complexity*, 4:67–382, 1994.

[14] Richard Beigel and John Gill. Counting classes: thresholds, parity, mods, and fewness. *Theoret. Comput. Sci.*, 103(1):3–23, 1992. 7th Annual Symposium on Theoretical Aspects of Computer Science (STACS 90) (Rouen, 1990).

[15] Richard Beigel and Jun Tarui. On ACC. In *32nd Annual Symposium on Foundations of Computer Science (San Juan, PR, 1991)*, pages 783–792. IEEE Comput. Soc. Press, Los Alamitos, CA, 1991.

[16] Amos Beimel, Yuval Ishai, Eyal Kushilevitz, and Jean-Francios Raymond. Breaking the $O\left(n^{1/(2k-1)}\right)$ barrier for information-theoretic private information retrieval. In *43rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 261–270, 2002.

[17] Avraham Ben-Aroya, Klim Efremenko, and Amnon Ta-Shma. Local list decoding with a constant number of queries. In *51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 715–722, 2010.

[18] E. Ben-Sasson and A. Gabizon. Extractors for polynomial sources over constant-size fields of small characteristic. Technical Report TR11-129, Electronic Colloquium on Computational Complexity, 2011.

[19] Eli Ben-Sasson and Swastik Kopparty. Affine dispersers from subspace polynomials. *SIAM J. Comput.*, 41(4):880–914, 2012.

[20] Eli Ben-Sasson, Shachar Lovett, and Noga Zewi. An additive combinatorics approach to the log-rank conjecture in communication complexity. *CoRR*, abs/1111.5884, 2011.

[21] Eli Ben-Sasson and Noga Zewi. From affine to two-source extractors via approximate duality. In *STOC*, pages 177–186, 2011.

[22] Vitaly Bergelson, Terence Tao, and Tamar Ziegler. An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}^\omega$. *Geom. Funct. Anal.*, 19(6):1539–1596, 2010.

[23] Arnab Bhattacharyya. Polynomial decompositions in polynomial time. In *Proc. 22nd Annual European Symposium on Algorithms*, pages 125–136, 2014.

[24] Arnab Bhattacharyya. Polynomial decompositions in polynomial time. In *Algorithms - ESA 2014 - 22th Annual European Symposium, Wroclaw, Poland, September 8-10, 2014. Proceedings*, pages 125–136, 2014.

[25] Arnab Bhattacharyya and Abhishek Bhowmick. Using higher-order fourier analysis over general fields. *CoRR*, abs/1505.00619, 2015.

[26] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *STOC*, pages 429–436, 2013.

[27] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *Proc. 45th Annual ACM Symposium on the Theory of Computing*, pages 429–436, 2013.

[28] Arnab Bhattacharyya, Eldar Fischer, and Shachar Lovett. Testing low complexity affine-invariant properties. In *Proc. 24th ACM-SIAM Symposium on Discrete Algorithms*, pages 1337–1355, 2013. `http://arxiv.org/abs/1201.0330v2`.

[29] Arnab Bhattacharyya, Elena Grigorescu, and Asaf Shapira. A unified framework for testing linear-invariant properties. In *Proc. 51st Annual IEEE Symposium on*

Foundations of Computer Science, pages 478–487, 2010.

[30] Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. Algorithmic regularity for polynomials and applications. In *Proc. 26th ACM-SIAM Symposium on Discrete Algorithms*, pages 1870–1889, 2015.

[31] Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. Algorithmic regularity for polynomials and applications. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1870–1889, 2015.

[32] Abhishek Bhowmick, Zeev Dvir, and Shachar Lovett. New bounds for matching vector families. *SIAM J. Comput.*, 43(5):1654–1683, 2014.

[33] Abhishek Bhowmick, Ariel Gabizon, Thái Hoàng Lê, and David Zuckerman. Deterministic extractors for additive sources: Extended abstract. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 277–286, 2015.

[34] Abhishek Bhowmick and Shachar Lovett. Nonclassical polynomials as a barrier to polynomial lower bounds. *CoRR*, abs/1412.4719, 2014. To appear, CCC 2015.

[35] Abhishek Bhowmick and Shachar Lovett. List decoding reed-muller codes over small fields. *CoRR*, abs/1407.3433, 2014. To appear, STOC 2015.

[36] Abhishek Bhowmick and Shachar Lovett. Bias vs structure of polynomials in large fields, and applications in effective algebraic geometry and coding theory. *CoRR*, abs/1506.02047, 2015.

[37] Abhishek Bhrushundi, Prahlad Harsha, and Srikanth Srinivasan. Aug, 2015. Personal Communication.

[38] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comp. Sys. Sci.*, 47:549–595, 1993. Earlier version in STOC'90.

[39] A. Bogdanov and E. Viola. Pseudorandom bits for polynomials. In *Proc.* 48^{th} *IEEE Symp. on Foundations of Computer Science (FOCS'07)*, 2007.

[40] Andrej Bogdanov. Pseudorandom generators for low degree polynomials. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 21–30, 2005.

[41] J. Bourgain. Estimates on exponential sums related to the diffiehellman distributions. *Geometric and Functional Analysis GAFA*, 15(1):1–34, 2005.

[42] J. Bourgain. On the construction of affine extractors. *Geometric and Functional Analysis*, 17:33–57, 2007.

[43] Jean Bourgain. Sumproduct theorems and exponential sum bounds in residue classes for general modulus. *Comptes Rendus Mathematique*, 344(6):349 – 352, 2007.

[44] Jean Bourgain, Zeev Dvir, and Ethan Leeman. Affine extractors over large fields with exponential error. *CoRR*, abs/1401.6189, 2014.

[45] W. Dale Brownawell. Bounds for the degrees in the nullstellensatz. *Annals of Mathematics*, 126(3):pp. 577–591, 1987.

[46] B. Buchberger. Ein algorithmus zum auffinden der basiselemente des restklassenrings nach einem nulldimensionalen polynomideal (an algorithm for finding the basis elements in the residue class ring modulo a zero dimensional polynomial ideal). In *PhD thesis, Universitt Innsbruck, 1965.* 1965.

[47] B. Chor, J. Friedman, O. Goldreich, J. Hastad, S. Rudich, and R. Smolensky. The bit extraction problem or t–resilient functions. In *Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 396–407, 1985.

[48] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988.

[49] Fan R. K. Chung, Persi Diaconis, and Ronald L. Graham. Random walks arising in random number generation. *Ann. Probab.*, 15(3):1148–1165, 07 1987.

[50] Ronald de Wolf. *A Brief Introduction to Fourier Analysis on the Boolean Cube.* Number 1 in Graduate Surveys. Theory of Computing Library, 2008.

[51] Pierre Deligne. La conjecture de Weil. I. *Inst. Hautes Études Sci. Publ. Math.*, (43):273–307, 1974.

[52] M. DeVos and A. Gabizon. Simple affine extractors using dimension expansion. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity*, 2010.

[53] Irit Dinur, Elena Grigorescu, Swastik Kopparty, and Madhu Sudan. Decodability of group homomorphisms beyond the johnson bound. In Richard E. Ladner and

Cynthia Dwork, editors, *STOC*, pages 275–284. ACM, 2008.

[54] Z. Dvir. Extractors for varieties. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity*, pages 102–113, 2009.

[55] Z. Dvir, A. Gabizon, and A. Wigderson. Extractors and rank extractors for polynomial sources. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 52–62, 2007.

[56] Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. *SIAM J. Comput.*, 40(4):1154–1178, 2011.

[57] Zeev Dvir and Guangda Hu. Matching-vector families and ldcs over large modulo. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:61, 2013.

[58] Klim Efremenko. 3-query locally decodable codes of subexponential length. In *Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing*, STOC '09, pages 39–44, New York, NY, USA, 2009. ACM.

[59] Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM J. Comput.*, 41(6):1694–1703, 2012.

[60] Peter Elias. List decoding for noisy channels. Research laboratory for electronics, MIT, 1957.

[61] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996.

[62] Alan M. Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[63] A. Gabizon and R. Raz. Deterministic extractors for affine sources over large fields. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 407–418, 2005.

[64] A. Gabizon and R. Shaltiel. Increasing the output length of zero-error dispersers. In Ashish Goel, Klaus Jansen, JosD.P. Rolim, and Ronitt Rubinfeld, editors, *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, volume 5171 of *Lecture Notes in Computer Science*, pages 430–443. Springer Berlin Heidelberg, 2008.

[65] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM J. Comput.*, 36(4):1072–1094, 2006.

[66] Joachim Von Zur Gathen, Marek Karpinski, and Igor Shparlinski. Counting curves and their projections. *Computational Complexity*, 6:64–99, 1996.

[67] O. Goldreich and L. Levin. A hard-core predicate for all one-way functions. In *Proc. 21^{st} ACM Symposium on the Theory of Computing*, pages 25–32, 1989.

[68] O. Goldreich, R. Rubinfeld, and M. Sudan. Learning polynomials with queries: The highly noisy case. *SIAM J. Discrete Math.*, 13(4):535–570, 2000.

[69] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45:653–750, 1998.

[70] Oded Goldreich, Howard Karloff, Leonard Schulman, and Luca Trevisan. Lower bounds for locally decodable codes and private information retrieval. In *17th IEEE Computational Complexity Conference (CCC)*, pages 175–183, 2002.

[71] Oded Goldreich and Tali Kaufman. Proximity oblivious testing and the role of invariances. In *Studies in Complexity and Cryptography*, pages 173–190. 2011.

[72] Oded Goldreich and Dana Ron. On proximity oblivious testing. *SIAM J. Comput.*, 40(2):534–566, 2011.

[73] P. Gopalan. A Fourier-analytic approach to Reed-Muller decoding. In *Proc.* 51^{st} *IEEE Symp. on Foundations of Computer Science (FOCS'10)*, pages 685–694, 2010.

[74] P. Gopalan, A. Klivans, and D. Zuckerman. List decoding Reed-Muller codes over small fields. In *Proc.* 40^{th} *ACM Symposium on the Theory of Computing (STOC'08)*, pages 265–274, 2008.

[75] Parikshit Gopalan, Venkatesan Guruswami, and Richard J. Lipton. Algorithms for modular counting of roots of multivariate polynomials. *Algorithmica*, 50(4):479–496, 2008.

[76] Parikshit Gopalan, Ryan O'Donnell, Rocco A. Servedio, Amir Shpilka, and Karl Wimmer. Testing Fourier dimensionality and sparsity. In *Proc. 36th Annual International Conference on Automata, Languages, and Programming*, pages 500–512, 2009.

[77] William T. Gowers. A new proof of Szeméredi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

[78] William T. Gowers. A new proof of Szeméredi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[79] William Timothy Gowers. A new proof of szemer'edis theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 17(2):230–261, 1998.

[80] B. Green and T. Tao. The distribution of polynomials over finite fields, with applications to the gowers norms. *Contrib. Discrete Math*, 4(2):1–36, 2009.

[81] Ben Green. Finite field models in additive combinatorics. In Bridget S Webb, editor, *Surveys in combinatorics 2005*, pages 1–27. Cambridge Univ. Press, 2005.

[82] Ben Green and Terence Tao. An inverse theorem for the Gowers U^3-norm. *Proc. Edin. Math. Soc.*, 51:73–153, 2008.

[83] Ben Green and Terence Tao. Linear equations in primes. *Ann. of Math.*, 171:1753–1850, 2010.

[84] Ben Green, Terence Tao, and Tamar Ziegler. An inverse theorem for the Gowers U^4-norm. *Glasgow Math. J.*, 53(1):1–50, 2011. `http://arxiv.org/abs/0911.5681`.

[85] Ben Green, Terence Tao, and Tamar Ziegler. An inverse theorem for the Gowers U^{s+1}-norm. *Ann. of Math.*, 176(2):1231–1372, 2012.

[86] Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20:71–86, 2000.

[87] Vince Grolmusz. Constructing set-systems with prescribed intersection sizes. *Journal of Algorithms*, 44:321–337, 2002.

[88] V. Guruswami. *List Decoding of Error-Correcting Codes*, volume 3282 of *Lecture Notes in Computer Science*. Springer, 2004.

[89] V. Guruswami. *Algorithmic Results in List Decoding*, volume 2 of *Foundations and Trends in Theoretical Computer Science*. Now Publishers, 2006.

[90] Venkat Guruswami and Madhu Sudan. Improved decoding of Reed-Solomon and algebraic-geometric codes. *IEEE Transactions on Information Theory*, 45:1757–1767, 1999.

[91] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. Unbalanced expanders and randomness extractors from parvaresh–vardy codes. *J. ACM*, 56(4):20:1–20:34, July 2009.

[92] R. W. Hamming. Error Detecting and Error Correcting Codes. *Bell System Techincal Journal*, 29:147–160, 1950.

[93] Johan Hastad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001.

[94] Hamed Hatami and Shachar Lovett. Estimating the distance from testable affine-invariant properties. In *Proc. 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 237–242. IEEE, 2013.

[95] Grete Hermann. Die frage der endlich vielen schritte in der theorie der polynomideale. *Mathematische Annalen*, 95(1):736–788, 1926.

[96] Bernard Host and Bryna Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math.*, 161(1):397–488, 2005.

[97] Ming-Deh Huang and Yiu-Chung Wong. An algorithm for approximate counting of points on algebraic sets over finite fields. In JoeP. Buhler, editor, *Algorithmic Number Theory*, volume 1423 of *Lecture Notes in Computer Science*, pages 514–527. Springer Berlin Heidelberg, 1998.

[98] Ming-Deh A. Huang and Doug Ierardi. Counting rational points on curves over finite fields (extended abstract). In *34th Annual Symposium on Foundations of Computer Science, Palo Alto, California, USA, 3-5 November 1993*, pages 616–625, 1993.

[99] Yuval Ishai and Eyal Kushilevitz. On the hardness of information-theoretic multiparty computation. In *Eurocrypt 2004*, volume 3027 of Lecture Notes in Computer Science, pages 439–455. Springer, Berlin, Heidelberg, 2004.

[100] Toshiya Itoh and Yasuhiro Suzuki. New constructions for query-efficient locally decodable codes of subexponential length. *IEICE Transactions on Information and Systems*, pages 263–270, 2010.

[101] J. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55:414–440, 1997.

[102] S. Johnson. A new upper bound for error-correcting codes. *Information Theory, IRE Transactions on*, 8(3):203–207, April 1962.

[103] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *Proc. 29th Annual IEEE Symposium on Foundations of Computer*

Science, pages 68–80, 1988.

[104] J. Kamp and D. Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM Journal on Computing*, 36:1231–1247, 2006.

[105] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *32nd ACM Symposium on Theory of Computing (STOC)*, pages 80–86, 2000.

[106] Nicholas Katz. On a theorem of ax. *American Journal of Mathematics*, 93:485–499, 1971.

[107] T. Kaufman, S. Lovett, and E. Porat. Weight distribution and list-decoding size of Reed-Muller codes. In *Innovations in Computer Science (ICS'10)*, pages 422–433, 2010.

[108] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 166–175, 2008.

[109] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. on Comput.*, 36(3):779–802, 2006.

[110] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 403–412, 2008.

[111] Kiran S. Kedlaya and Sergey Yekhanin. Locally decodable codes from nice subsets of finite fields and prime factors of Mersenne numbers. *SIAM Journal on Computing*, 38:1952–1969, 2009.

[112] Iordanis Kerenidis and Ronald de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *Journal of Computer and System Sciences*, 69:395–420, 2004.

[113] Janos Kollar. Sharp effective nullstellensatz. *Journal of the American Mathematical Society*, 1(4):pp. 963–975, 1988.

[114] D. Koller and N. Megiddo. Constructing small sample spaces satisfying given constraints. In *Proc. 25th Annual ACM Symposium on the Theory of Computing*, pages 268–277, 1993.

[115] Swastik Kopparty and Sergey Yekhanin. Detecting rational points on hypersurfaces over finite fields. In *Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*, CCC '08, pages 311–320, Washington, DC, USA, 2008. IEEE Computer Society.

[116] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *SIAM Journal of Computing*, 22(6):1331–1348, 1993.

[117] X. Li. A new approach to affine extractors and dispersers. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity*, 2011.

[118] Xin Li. Extractors for affine sources with polylogarithmic entropy. Technical report, ECCC, 2015. `http://eccc.hpi-web.de/report/2015/121/`.

[119] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Cambridge University Press, Cambridge, 1983.

[120] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(3):69–82, 2009.

[121] Shachar Lovett. Holes in generalized reed-muller codes. *IEEE Transactions on Information Theory*, 56(6):2583–2586, 2010.

[122] Shachar Lovett. An exposition of Sanders' quasi-polynomial freiamn-ruzsa theorem. To appear., 2012.

[123] Edouard Lucas. Thorie des fonctions numriques simplement priodiques. *American Journal of Mathematics*, 1(2):pp. 184–196, 1878.

[124] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error Correcting Codes*. North Holland, Amsterdam, New York, 1977.

[125] Yeow Meng Chee, Tao Feng, San Ling, Huaxiong Wang, and Liangfeng Zhang. Query-efficient locally decodable codes of subexponential length. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR10-173, 2010.

[126] Elchanan Mossel, Ryan ODonnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Ann. of Math.*, 171(1), 2010.

[127] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993.

[128] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. on Comput.*, 22(4):838–856, 1993. Earlier version in STOC'90.

[129] Michael Navon and Alex Samorodnitsky. On delsarte's linear programming bounds for binary codes. In *Proc. 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 327–338, 2005.

[130] N. Nisan and A. Wigderson. Hardness vs. randomness. *Journal of Computer and System Sciences*, 49:149–167, 1994.

[131] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1):43–52, 1996.

[132] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[133] R. Pellikaan and X. Wu. List decoding of q-ary Reed-Muller codes. *IEEE Transactions on Information Theory*, 50(4):679–682, 2004.

[134] Prasad Raghavendra. A note on Yekhanin's locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR07-016, 2007.

[135] A. Rao. Extractors for a constant number of polynomially small min-entropy independent sources. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 497–506, 2006.

[136] A. Rao. An exposition of Bourgain's 2-source extractor. Technical Report TR07-034, Electronic Colloquium on Computational Complexity, 2007.

[137] Anup Rao. Extractors for low-weight affine sources. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(015), 2008.

[138] A. A. Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki*, 41(4):598–607, 623, 1987.

[139] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. on Comput.*, 25:252–271, 1996.

[140] By Imre Ruzsa and Imre Z. Ruzsa. An analog of Freiman's theorem in groups, 1993.

[141] Imre Z. Ruzsa. Sumsets and structure. In *Combinatorial number theory and additive group theory*, Adv. Courses Math. CRM Barcelona, pages 87–210. Birkhäuser Verlag, Basel, 2009.

[142] T. Sanders. On the Bogolyubov-Ruzsa lemma. *ArXiv e-prints*, October 2010.

[143] M. Santha and U. V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33:75–87, 1986.

[144] W.M. Schmidt. *Equations over Finite Fields. An Elementary Approach*, volume 536 of *Lecture Notes in Mathematics*. Springer-Verlag, 1976.

[145] Arnold Schönhage and Volker Strassen. Schnelle multiplikation grosser zahlen. *Computing*, 7:281–292, 1971.

[146] J. T. Schwartz. Fast probabilistic algorithms for verication of polynomial identities. *Journal of the ACM*, 27:701–717, 1980.

[147] Jiri Sgall. Bounds on pairs of families with restricted intersections. *Combinatorica*, 19:555–566, 1999.

[148] Ronen Shaltiel. Dispersers for affine sources with sub-polynomial entropy. In *FOCS*, pages 247–256, 2011.

[149] Ronen Shaltiel and Christopher Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *J. ACM*, 52(2):172–216, 2005.

[150] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001.

[151] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proc.* 19^{th} *ACM Symposium on Theory of Computing (STOC'87)*, pages 77–82, 1987.

[152] Daniel Štefankovič. Fourier transform in computer science. Master's thesis, University of Chicago, 2000.

[153] M. Sudan. Decoding of Reed-Solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180–193, 1997.

[154] M. Sudan. List decoding: Algorithms and applications. *SIGACT News*, 31(1):16–27, 2000.

[155] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. In *39th ACM Symposium on Theory of Computing (STOC)*, pages 537–546, 1999.

[156] Endre Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[157] A. Ta-Shma, D. Zuckerman, and S. Safra. Extractors from Reed-Muller codes. In *Proc. 42^{nd} IEEE Symp. on Foundations of Computer Science (FOCS'01)*, pages 638–647, 2001.

[158] T. Tao and V. Vu. *Additive Combinatorics*. Cambridge University Press, 2006.

[159] T. Tao and T. Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *ArXiv e-prints*, January 2011.

[160] T. Tao and T. Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *ArXiv e-prints*, January 2011.

[161] Terence Tao. *Higher Order Fourier Analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012.

[162] Terence Tao. Expanding polynomials over finite fields of large characteristic, and a regularity lemma for definable sets. http://arxiv.org/pdf/1211.2894v4.pdf , 2013.

[163] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Analysis & PDE*, 3(1):1–20, 2010.

[164] Terence Tao and Tamar Ziegler. The inverse conjecture for the Gowers norm over finite fields in low characteristic. *Ann. Comb.*, 16(1):121–188, 2012.

[165] L. Trevisan. List-decoding using the XOR lemma. In *Proc. 44^{th} IEEE Symposium on Foundations of Computer Science (FOCS'03)*, page 126, 2003.

[166] Luca Trevisan. Some applications of coding theory in computational complexity. *Quaderni di Matematica*, 13:347–424, 2004.

[167] Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *Computational Complexity, 2009. CCC'09. 24th Annual IEEE Conference on*, pages 126–136. IEEE, 2009.

[168] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.

[169] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Computational Complexity*, 18(2):209–217, 2009.

[170] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Proc. 22nd Annual IEEE Conference on Computational Complexity*, pages 141–154, June 2007.

[171] A. Weil. On some exponential sums. *Proceedings of the National Academy of Sciences*, 34:204–207, 1948.

[172] André Weil. *Sur les courbes algébriques et les variétés qui s'en déduisent.* Actualités Sci. Ind., no. 1041 = Publ. Inst. Math. Univ. Strasbourg **7** (1945). Hermann et Cie., Paris, 1948.

[173] Hermann Weyl. Über ein problem aus dem gebiete der diophantischen approximationen. *Nachr. Ges. Wiss. Gttingen*, pages 234–244, 1914.

[174] David Woodruff. New lower bounds for general locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR07-006, 2007.

[175] David P. Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. In *Proceedings of the 13th international conference on Approximation, and 14 the International conference on Randomization, and combinatorial optimization: algorithms and techniques*, APPROX/RANDOM'10, pages 766–779, Berlin, Heidelberg, 2010. Springer-Verlag.

[176] J. Wozencraft. List decoding. Technical Report 48:90-95, Quarterly Progress Report, Research Laboratory of Electronics, MIT, 1958.

[177] Amir Yehudayoff. Affine extractors over prime fields. *Combinatorica*, 31(2):245–256, 2011.

[178] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55:1–16, 2008.

[179] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *J. ACM*, 55(1):1:1–1:16, February 2008.

[180] Sergey Yekhanin. Locally decodable codes. *Foundations and Trends in Theoretical Computer Science*, 6(3):139–255, 2012.

[181] Yuichi Yoshida. A characterization of locally testable affine-invariant properties via decomposition theorems. In *Proc. 46th Annual ACM Symposium on the Theory of Computing*, pages 154–163, 2014.

[182] Chen Yuan, Qian Guo, and Haibin Kan. A novel elementary construction of matching vectors. *Information Processing Letters*, 112(12):494 – 496, 2012.

[183] R E. Zippel. Probabilistic algorithms for sparse polynomials. *Proceedings of EUROSAM*, pages 216–226, 1979.

Vita

Abhishek Bhowmick received the Bachelor of Technology degree in Computer Science and Engineering from the Indian Institute of Technology (IIT) Kanpur in May 2010. He is currently pursuing a Ph.D. in Computer Science under the supervision of Prof. David Zuckerman at The University of Texas at Austin.

Permanent address: 3543 Greystone Drive, Austin TX-78731.

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.